THE PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

Published by the American Institute of Physics for the

AMERICAN PHYSICAL SOCIETY

Officers

President

R. T. Birge
University of California
Berkeley, California

Vice-President

R. P. Wiener
Princeton University
Princeton, New Jersey

Secretary

K. K. Darrow
Bell Telephone Laboratories
Office:
Columbia University
New York 27, New York

Treasurer

G. B. Pegram
Columbia University
New York 27, New York

Local Secretary for the Pacific Coast

William A. Nierenberg
University of California
Berkeley, California

Managing Editor

S. A. Goudsmit
Brookhaven National Laboratory
Upton, Long Island
New York

BOARD OF EDITORS

Editor

S. A. Goudsmit

Assistant Editor

S. Pasternack

Associate Editors

1953–1955	1954–1956	1955–1957
W. A. Fowler	J. G. Beckerley	G. J. Dienes
Kenneth Greisen	Gregory Breit	R. G. Herb
Charles Kittel	F. J. Dyson	K. G. McKay
R. E. Marshak	P. Kusch	J. R. Pellam
W. K. H. Panofsky	R. Sagdeev	Lewi Tonks
V. F. Weisskopf	J. C. Street	Gregory Wannier

Manuscripts for publication should be submitted to S. A. Goudsmit, Brookhaven National Laboratory, Upton, Long Island, New York. The authors' institutions are requested to pay a publication charge of $25 per page which, if honored, entitles them to 100 free reprints. Instructions will be sent with galley proofs.

Proof and all correspondence concerning papers in the process of publication should be addressed to the Publication Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price

United States and Canada.....$30.00
Elsewhere.....$32.00

Back Numbers[1]

Yearly back number rate when complete year is available:
Prior to 1949: $25.00
Thereafter: $30.00
Single copies: $3.00 each, prior to July, 1929;
$1.50 each, July, 1929–December, 1941;
$3.00 each, January, 1942–December, 1945;
$1.50 each, thereafter.
General index, 1893–1920: $6.00.
1921–1950: Paper binding, $20 (Members); $24 (Nonmembers)
Cloth binding, $12 (Members); $16 (Nonmembers)

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Changes of address in the case of members of the American Physical Society should be addressed to the Treasurer; in the case of other subscribers to the American Institute of Physics.

[1] Bur Series 1 Inquire of Physics Department, Central Publication, Ithaca, New York.

The Physical Review is published semi-monthly at Prince and Lemon Streets, Lancaster, Pennsylvania.

Entered as second-class matter March 6, 1917, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.

Acceptance for mailing at special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized January 31, 1925.

THE
PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 99, NO. 4 AUGUST 15, 1955

"Virial Theorem" for the Flow of Energy

R. EISENSCHITZ
Queen Mary College, University of London, London, England
(Received March 14, 1955)

Expressions are obtained for the flow of energy and momentum in assemblies of densely packed molecules, using a generalized virial theorem. They differ to some extent from expressions put forward by previous authors who had only derived the divergence of the flow vectors. Classical results are shown to be valid also in quantum mechanics.

INTRODUCTION

USES of the virial theorem in modern physics were recently published by Parker[1]; in the present paper, the theorem is applied to the molecular theory of irreversible processes.

Applying the conservation laws of mechanics, Born and Green[2] and Irving and Kirkwood[3] obtained expressions for the divergence of the flow of energy and momentum in fluids; their results remain valid in quantum theory as shown by Irving and Zwanzig.[4]

The same authors also gave expressions for the flow of energy which, however, do not agree with each other. They are, in fact, plausible but arbitrary conjectures, the flow being not uniquely determined by its divergence. Their expressions for the flow of momentum do agree and are in accordance with previous work by Falkenhagen.[5]

As these flow vectors are significant in many contexts, to quote only the flow properties of liquid helium and the sound absorption in dense gases, the subject is considered in the present paper again and from an alternative point of view. Instead of attempting to derive the fluxes by the way of their divergence they are obtained by applying a generalized virial theorem that is due to Rayleigh[6]; he (like Clausius before him) obtained the virial theorem in the form of a relation between time averages. Its validity is then restricted to periodic movement; in quantum mechanics the corresponding theorem applies only to steady pure states. In the present paper, results are freed from these restrictions by obtaining relations between ensemble averages.

PREMISES

The macroscopic properties of an assembly of molecules in a container are assumed to be determined by a probability distribution in phase space which, if not independent of the time, has a stationary projection on two-molecule subspaces.

Forces between molecules are derived from a potential energy depending on the distances between the molecules

$$\mathrm{X}=\tfrac{1}{2}\sum_j\sum_{k\neq j}\chi_{jk}, \tag{1}$$

where χ_{jk} is the familiar potential energy of interaction between isolated molecules ($j, k=1, 2 \ldots N$). Forces arising from the interaction of molecules with the wall of the container are derived from a potential energy

$$\Omega=\sum_j \omega_j, \tag{2}$$

which depends partly on dynamical variables appropriate to the wall, such as the position of the molecules of the container.

Movement of the liquid molecules is determined by equations of motion

$$\begin{aligned} d\mathbf{r}_j/dt&=\mathbf{p}_j/m,\\ d\mathbf{p}_j/dt&=\sum_{k\neq j}\mathbf{f}_{jk}+\mathbf{F}_j, \end{aligned} \tag{3}$$

[1] E. N. Parker, Phys. Rev. **96**, 1686 (1954).
[2] M. Born and H. S. Green, Proc. Roy. Soc. (London) **A190**, 455 (1947).
[3] J. H. Irving and J. G. Kirkwood, J. Chem. Phys. **18**, 817 (1950).
[4] J. H. Irving and W. Zwanzig, J. Chem. Phys. **19**, 1173 (1951).
[5] H. Falkenhagen, Physik. Z. **32**, 745 (1931).
[6] W. J. Rayleigh, Phil. Mag. **50**, 210 (1900).

where $\mathbf{F}_j$ and $\mathbf{f}_{jk}=-\mathbf{f}_{kj}$ are the forces on the molecule j arising from the potentials X and Ω, respectively, and $\mathbf{r}_j$ is its position; m is the molecular mass.

The change in time of the "wall variables" is not specified in detail. It is postulated that certain averages of functions of these variables are stationary and can be expressed in terms of macroscopic quantities. In particular, the average of the forces $\mathbf{F}_j$ is supposed to balance external forces applied at the surface in form of stresses also the average work done by these forces is assumed to be equal to the energy as supplied from external sources through the unity of surface area per unity of time (denoted by $-W$). In quantitative terms

$$\langle\sum_j \mathbf{F}_j\,\delta(\mathbf{r}_j-\mathbf{r})\rangle=-\mathbf{n}\cdot\mathbf{P}dS, \tag{4}$$

$$(1/m)\langle\sum_j \mathbf{F}_j\cdot\mathbf{p}_j\,\delta(\mathbf{r}_j-\mathbf{r})\rangle=-\mathbf{n}\cdot\mathbf{Q}dS, \tag{5}$$

where δ is the delta function, $\mathbf{P}$ the stress tensor, dS and $\mathbf{n}$ the surface element and $-$normal, $\mathbf{Q}$ the flow of energy ($W=\mathbf{n}\cdot\mathbf{Q}$) and $\mathbf{r}$ the position vector in 3-dimensional space. It follows that

$$\langle\sum_j \mathbf{r}_j\mathbf{F}_j\rangle=-\int \mathbf{rn}\cdot\mathbf{Q}dS, \tag{6}$$

$$(1/m)\langle\sum_j \mathbf{r}_j\mathbf{F}_j\cdot\mathbf{p}_j\rangle=-\int \mathbf{rn}\cdot\mathbf{Q}dS. \tag{7}$$

The surface integrals are evaluated by means of the Gaussian theorem

$$\int \mathbf{rn}\cdot\mathbf{j}dS=\int \mathbf{j}dV,$$

where $\mathbf{j}$ is a vector or tensor subject to the condition div $\mathbf{j}=0$ and V is the volume. If $\mathbf{P}$ and $\mathbf{Q}$ are constant it follows that

$$\int \mathbf{rn}\cdot\mathbf{P}dS=V\mathbf{P}, \tag{8}$$

$$\int \mathbf{rn}\cdot\mathbf{Q}dS=V\mathbf{Q}. \tag{9}$$

The above assumptions regarding the forces between the molecules and the wall must be valid unless the stress is balanced by any alternative forces or if the work done by these forces is maintained by an energy supply alternative to that which is taken into account.

Let $A, B \ldots$ be functions of the dynamical variables in the form of a sum, each term of which depends on the coordinates and momenta of not more than two molecules. Then, defining dA/dt in accordance with Eqs. (3), on account of the assumption of stationary conditions:

$$\langle dA/dt\rangle=\langle dB/dt\rangle=\cdots=0. \tag{10}$$

By choosing appropriate expressions for $A, B \ldots$ relations between physically important ensemble averages are obtained.

FLOW OF MOMENTUM

Let

$$A=\sum_j \mathbf{r}_j\mathbf{p}_j, \tag{11}$$

so that

$$dA/dt=\sum_j[(\mathbf{p}_j\mathbf{p}_j/m)+\mathbf{r}_j\mathbf{F}_j+\sum_{k\neq j}\mathbf{r}_j\mathbf{f}_{jk}]. \tag{12}$$

The last term may be written as

$$\tfrac{1}{2}\sum_j\sum_{k\neq j}\mathbf{f}_{jk}(\mathbf{r}_j-\mathbf{r}_k).$$

The average of the second term is, for constant $\mathbf{P}$, according to Eq. (8), equal to $-V\mathbf{P}$. By averaging Eq. (12) over the ensemble of systems it follows that

$$\mathbf{P}=(N/Vm)\langle\mathbf{p}_j\mathbf{p}_j\rangle+(N^2/2V)\langle(\mathbf{r}_j-\mathbf{r}_k)\mathbf{f}_{jk}\rangle, \tag{13}$$

where N is the total number of molecules and the subscripts j, k refer to a representative pair. Equation (13) agrees with the results obtained by previous authors.

FLOW OF ENERGY

Let

$$B=(1/2m)\sum_j p_j^2\mathbf{r}_j, \tag{14}$$

so that

$$dB/dt=\sum_j\Big[(p_j^2\mathbf{p}_j/2m^2)+(\mathbf{r}_j\mathbf{p}_j\cdot\mathbf{F}_j/m)+(1/m)\sum_{k\neq j}\mathbf{r}_j\mathbf{p}_j\cdot\mathbf{f}_{jk}\Big]. \tag{15}$$

The last term may be written as

$$(1/2m)\sum_j\sum_{k\neq j}(\mathbf{r}_j\mathbf{p}_j-\mathbf{r}_k\mathbf{p}_k)\cdot\mathbf{f}_{jk}$$

or

$$\sum_j\sum_{k\neq j}(1/4m)[(\mathbf{r}_j+\mathbf{r}_k)(\mathbf{p}_j-\mathbf{p}_k)+(\mathbf{r}_j-\mathbf{r}_k)(\mathbf{p}_j+\mathbf{p}_k)]\cdot\mathbf{f}_{jk}.$$

The average of the second term is, for constant $\mathbf{Q}$, according to Eq. (9), equal to $-V\mathbf{Q}$.

By averaging Eq. (15) over the ensemble of systems it follows that

$$Q=(N/2Vm^2)\langle p_j^2\mathbf{p}_j\rangle+(N^2/4Vm)\langle[(\mathbf{r}_j+\mathbf{r}_k)(\mathbf{p}_j-\mathbf{p}_k)+(\mathbf{r}_j-\mathbf{r}_k)(\mathbf{p}_j+\mathbf{p}_k)]\cdot\mathbf{f}_{jk}\rangle, \tag{15a}$$

where the subscripts j and k refer to a representative pair of molecules. This expression will be transformed in order to make it comparable with the expressions obtained by previous authors. Let

$$C=\sum_j\sum_{k\neq j}(\mathbf{r}_j+\mathbf{r}_k)\chi_{jk}. \tag{16}$$

As

$$m d\chi_{jk}/dt=(\partial\chi_{jk}/\partial\mathbf{r}_j)\cdot\mathbf{p}_j+(\partial\chi_{jk}/\partial\mathbf{r}_k)\cdot\mathbf{p}_k=-\mathbf{f}_{jk}\cdot(\mathbf{p}_j-\mathbf{p}_k),$$

it follows that

$$dC/dt=(1/m)\sum_j\sum_{k\neq j}[\chi_{jk}(\mathbf{p}_j+\mathbf{p}_k) - (\mathbf{r}_j+\mathbf{r}_k)\mathbf{f}_{jk}\cdot(\mathbf{p}_j-\mathbf{p}_k)]. \quad (17)$$

By taking the average of Eq. (17) over the ensemble of systems, the relation

$$\langle(\mathbf{r}_j+\mathbf{r}_k)\mathbf{f}_{jk}\cdot(\mathbf{p}_j-\mathbf{p}_k)\rangle=\langle\chi_{jk}(\mathbf{p}_j+\mathbf{p}_k)\rangle$$

is obtained; the subscripts j and k refer to a representative pair of molecules. This relation is used for transforming the second term of Eq. (15a). Hence,

$$\mathbf{Q}=(N/2Vm^2)\langle p_j^2\mathbf{p}_j\rangle + (N^2/4Vm)\langle[\chi_{jk}+(\mathbf{r}_j-\mathbf{r}_k)\mathbf{f}_{jk}\cdot](\mathbf{p}_j+\mathbf{p}_k)\rangle. \quad (18)$$

Formulas given by previous authors do not fully agree with Eq. (18). Born and Green[2] [in their Eqs. (2.17) and (4.15)] give an expression in which the term depending on χ_{jk} in (18) is missing. Irving and Kirkwood[3] (in their Eqs. (6.21) and (6.24)) give an expression in which the factor $(\mathbf{p}_j+\mathbf{p}_k)$ of Eq. (18) is replaced by $2\mathbf{p}_j$ which is equal to $(\mathbf{p}_j+\mathbf{p}_k)+(\mathbf{p}_j-\mathbf{p}_k)$. Unless the terms by which the various expressions for $\mathbf{Q}$ differ make no appreciable contribution to the statistical average, the flow of energy, in particular the heat flow in a steady temperature gradient, should be derived from (18) rather than from the older equations.

QUANTUM THEORY

Macroscopic properties of the assembly are now determined by a probability distribution for the occurrence of pure states in a mixture. The states are chosen in such a way that they are steady with respect to N-2 molecules. The distribution, in general time dependent, is in the present case assumed to permit averaging with respect to the states of N-2 molecules and to yield an average that is independent of the time.

Equations (3) and (10) remain valid, differentiations with respect to the time being expressed in terms of the appropriate Poisson bracket, such as $(i/\hbar)(H\mathbf{r}_k-\mathbf{r}_kH)$ for $d\mathbf{r}_k/dt$. The Hamilton function H is defined as

$$H=\sum_j p_j^2/2m+\Omega+\mathrm{X}.$$

Equations (4) and (5) must not be used explicitly since they apply to a molecule that is localized at some point near the wall. Equations (6) and (7) which are in classical theory consequences of (4) and (5) should now be valid on their own merit expressing the balance of forces and energy; Eq. (7) will, however, have to be made symmetric with respect to commutations and, accordingly replaced by

$$(1/2m)\langle\sum_j(\mathbf{r}_j\mathbf{F}_j\cdot\mathbf{p}_j+\mathbf{p}_j\cdot\mathbf{F}_j\mathbf{r}_j)\rangle=-\int\mathbf{rn}\cdot\mathbf{Q}dS. \quad (19)$$

The functions A, B are defined also in a symmetric way

$$A=\tfrac{1}{2}\sum_j(\mathbf{r}_j\mathbf{p}_j+\mathbf{p}_j\mathbf{r}_j), \quad (20)$$

$$B=(1/2m)\sum_j(\mathbf{r}_jp_j^2+p_j^2\mathbf{r}_j), \quad (21a)$$

or, alternatively,

$$B=(1/2m)\sum_j(\mathbf{p}_j\cdot-\mathbf{r}_j-\cdot\mathbf{p}_j), \quad (21b)$$

the dots and hyphens signifying scalar multiplication of the first and third factor.

In calculating dA/dt, etc., the following commutation rules are used:

$$uG(x)-G(x)u=(\hbar/i)(dG/dx),$$

$$G(u)x-xG(u)=(\hbar/i)(dG/du),$$

$$u^2G(x)-G(x)u^2=(\hbar/i)[(dG/dx)u+u(dG/dx)].$$

Here x and u denote a Cartesian coordinate and the conjugate component of momentum. $G(x)$ and $G(u)$ are functions of these variables.

By averaging the time derivatives dA/dt, dB/dt, dC/dt over the ensemble of systems expressions for $\mathbf{P}$ and $\mathbf{Q}$ are found which are identical with those presented in Eqs. (13) and (18), respectively. These expressions do not contain any noncommuting factors. Equation (13) applies equally to normal and tangential stress components. Equation (18) is arrived at whether B is defined according to (21a) or according to (21b). The noncommuting factors in B do not give rise to any ambiguity. Averages that are indicated in (13) or (18) can be evaluated by the method of Irving and Zwanzig[4] or by means of the density matrix.

CONCLUSION

Results obtained in this paper do not depend on any arbitrary assumption. They apply to steady conditions; application to time dependent processes is possible if the left-hand sides of Eqs. (12), (15), and (17) are taken into account. In periodic processes, such as the absorption of sound, a time average of the ensemble averages may be taken; if this reduces the mean time derivatives (dA/dt, dB/dt, dC/dt) to zero, Eqs. (13) and (18) remain valid in their present form.

Behavior of a Bose System

G. V. CHESTER
*Yale University, New Haven, Connecticut**
(Received March 25, 1955)

It is shown that a weakly interacting system of Bose particles undergoes a second-order transition both under conditions of constant volume and under conditions of constant pressure. This type of behavior is qualitatively identical with that exhibited by liquid helium.

IT is now well established[1,2] that a system composed of noninteracting particles obeying Bose-Einstein statistics undergoes a phase transition. If the density of the system is taken to be equal to that of liquid helium and if the mass of the particles is taken to be that of a helium atom, then the transition takes place at about 3°K. This fact has led to many attempts to relate the λ transition that occurs in liquid helium, at about 2°K, to the transition exhibited by a system of noninteracting Bose particles. While it is quite clear that such a system cannot accurately account for the properties of a liquid, it is, nevertheless, true that theories based on it have had some success in correlating the experimental properties of liquid helium. In this note we shall compare the transition in the liquid with the transition that occurs in a system of *weakly interacting* Bose particles. It must be emphasized that in this type of system we can only take into account weak attractive and repulsive forces between the particles; we are quite unable to deal with the very strong repulsive forces that ultimately come into play when the helium atoms approach sufficiently close to one another. However, this type of system has the advantage that it can be given a fairly rigorous treatment by means of perturbation theory.[3]

The transition in liquid helium appears to be a second-order transition whether we study it under conditions of constant volume or under conditions of constant pressure. For instance, the specific heat curves at constant pressure and constant volume are very similar in shape and both exhibit the same type of discontinuity at the λ point. The transition that occurs in a system of noninteracting Bose particles is, however, a first-order transition if we study the system under conditions of constant pressure, and a third-order transition if we study it under conditions of constant volume. Thus, although such a system does exhibit a transition, the details of it are very different from those of the λ transition in the liquid. This fact has been emphasized by Rice[4] and London.[5]

In I the author discussed some of the properties of a system of weakly interacting Bose particles. It was shown that such a system undergoes a second-order transition under conditions of constant volume. We shall now show that it undergoes a second-order transition under conditions of constant pressure. The free energy of the system is, to the accuracy of first-order perturbation theory,

$$F=F_0+g(F_1+F_1'), \tag{1}$$

where F_0 is the free energy of the noninteracting system, g is the coupling constant between the particles, F_1 and F_1' are given by[3]

$$\left.\begin{aligned} F_1&=\tfrac{1}{2}\sum_{\mathbf{k}\neq\mathbf{k}'}\sum \bar{n}_{\mathbf{k}}\bar{n}_{\mathbf{k}'}K_{\mathbf{k},\mathbf{k}'}+\bar{n}_0\sum_{\mathbf{k}\neq 0}\bar{n}_{\mathbf{k}}K_{\mathbf{k},0},\\ F_1'&=\tfrac{1}{2}N(N-1)K_{0,0}. \end{aligned}\right\} \tag{2}$$

These equations follow at once from Eqs. (2.15) and (1.2) of I. It is easily seen from Eq. (2.14) of I that $K_{\mathbf{k},\mathbf{k}'}$ is proportional to V^{-1}, for all $\mathbf{k}$ and $\mathbf{k}'$. Since each summation over the wave vectors $\mathbf{k}$ introduces a factor proportional to V the first term in F_1 is proportional to V. On the other hand, $\bar{n}_0$ is equal to $N[1-C(T)V]$, where $C(T)$ is a function of T alone. The second term in F_1 therefore consists of a term that is proportional to V and a term that is independent of V. Equation (2.27) of I shows in fact that F_1 is of the form

$$F_1=N[a(T)\rho^{-1}+b(T)]; \tag{3}$$

while from Eq. (2.14) of I we see that F_1' is given by

$$F_1'=2\pi N\rho\int_{2r_0}^{\infty} w(r)r^2dr, \tag{4}$$

where $a(T)$ and $b(T)$ are functions of the temperature alone, $w(r)$ is the interaction potential between the particles and r_0 is a cut-off parameter. We first note that $(\partial^2F_0/\partial V^2)_T\equiv 0$, $T\leq T_\lambda$; next, from Eqs. (3) and (4) we see that $(\partial^2F_1/\partial V^2)_T\equiv 0$, while $(\partial^2F_1'/\partial V^2)_T>0$ provided that

$$\int_{2r_0}^{\infty} w(r)r^2dr>0. \tag{5}$$

If we take the same Lennard-Jones (6, 12) potential that we used in I, *viz.* Eq. (4.1) of I, then we find that the inequality (5) will be satisfied as long as $r_0<1.14$ A.

* Contribution No. 1282 from the Sterling Chemistry Laboratory.

[1] A. R. Fraser, Phil. Mag. **42**, 156, 165 (1951).
[2] R. H. Fowler and H. Jones, Proc. Cambridge Phil. Soc. **34**, 573 (1938).
[3] G. V. Chester, Phys. Rev. **94**, 246 (1954). Hereafter referred to as I.
[4] O. K. Rice, Phys. Rev. **93**, 1161 (1954).
[5] F. London, *Superfluids* (John Wiley and Sons, Inc., New York, 1954), Vol. 2, p. 52.

The value of r_0 used in I was 1.09 A, so the inequality is easily satisfied. Consequently $(\partial p/\partial V)_T<0$, for $T<T_\lambda$, and the system is stable below the λ temperature. If we bear in mind that $(\partial p/\partial V)_T$ is discontinuous at T_λ then this implies that the system undergoes a second-order transition under conditions of constant pressure. The inequality (5) has a simple physical interpretation. It can only be satisfied if the potential is positive for some range of values of r. This simply means that it is the positive or "repulsive" part of the potential that prevents the particles from "condensing" into zero volume and exhibiting a first-order transition.

We can therefore conclude that a system of weakly interacting Bose particles undergoes a second-order transition both under conditions of constant pressure and constant volume. This behavior is qualitatively identical with that shown by liquid helium. These conclusions are, however, based on a treatment of a weakly interacting system and it remains to be shown that they carry over to the physically interesting case of a strongly interacting one.

The author wishes to thank the United States Government for the award of a Foreign Operations Administration research fellowship, during the tenure of which this work was carried out.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Quantum Effects in the Interaction between Free Electrons and Electromagnetic Fields

P. S. Faragó, *Department of Electromagnetic Waves, Central Research Institute for Physics, Budapest, Hungary*

AND

G. Marx, *Department of Theoretical Physics, Physical Institute of the Eötvös University, Budapest, Hungary*

(Received December 7, 1954)

An electron beam shot through a transverse rf field may suffer a directional spread. If experimental conditions are suitably chosen, the directional spread may be due only to the quantum dispersion of energy exchange between free electrons and rf field. A simple collector electrode system might allow not only the detection of the directional spread of the electrons, but the presence of a quantum effect might be checked by plotting the collector current *versus* rf field amplitude, the plot for the quantum effect being different from those for classical effects. The results of various theoretical treatments of the effect are briefly compared, both from the point of view of their principal foundations and of the possibility of their experimental verification.

In recent years several papers[1–4] have been published dealing with that part of the dispersion in energy exchange between free electrons and rf fields which the classical theory cannot predict. The problem has been treated from various points of view and differing results were obtained for the energy spread due to the quantum nature of the interaction.

One method[1,3] consists essentially in the assumption of the independence of the elementary emission and absorption processes, so that statistical methods can be applied in treating them. In this case the magnitude of quantum dispersion of the velocity of the electrons can be estimated by taking the classical energy gain W_+ and energy loss W_- during the accelerating and decelerating periods and interpreting them as the average absorption and emission of $N_+=W_+/\hbar\omega$ and $N_-=W_-/\hbar\omega$ photons respectively, with a standard deviation of

$$\Delta W=\hbar\omega(N_++N_-)^{\frac{1}{2}}. \tag{1}$$

Shulman[3] reports that he has found this effect by measuring the energy distribution of electrons shot through a longitudinal rf field in a wave guide. The difficulty of any experiment of this kind lies in the fact that the expected effect of quantum-mechanical origin and the classical energy spread resulting from the emission velocity distribution and from the dependence of transit time on the entrance phase angle of the electrons are of the same order of magnitude. The classical effects, however, can be made negligible in comparison with the quantum effect, if the electrons are shot through a transverse rf field and the transit angle of the electrons is properly chosen. In this case there should be a dispersion in the direction of the electron velocities caused only by the quantum nature of the energy exchange.

For the sake of simplicity we considered[5] a rectangular cavity excited in its TE_{012} mode. The only component of the electric field is, in this case:

$$E_z=E_0\sin(2\pi x/a)\cdot\sin(2\pi y/b)\cdot\sin\omega t,$$

the symbols having their usual meaning. For a given resonance frequency one can always find such a cavity that an electron entering along the x-direction with a given velocity v_0 spends an integral number of periods

[1] L. P. Smith, Phys. Rev. **69**, 195 (1946).
[2] J. C. Ward, Phys. Rev. **80**, 119 (1950).
[3] C. Shulman, Phys. Rev. **82**, 116 (1951); **83**, 4 (1951).
[4] I. R. Senitzky, Phys. Rev. **86**, 595 (1952); **90**, 386 (1953).
[5] P. S. Faragó and G. Marx, Acta Phys. Hung. **4**, 23 (1954).

k within it. It can be easily shown that such an electron (independently of its entrance phase angle) leaves the cavity without changing its direction, and the parallel shift of the beam cannot be detected for reasonable rf amplitudes and frequencies above 10 kMc/sec. If the initial velocity of the electrons is $v=v_0(1\pm\epsilon)$ ($\epsilon\ll 1$), a directional spread of the order of magnitude

$$\Delta\varphi_c\sim(2\pi eE_0/v_0m\omega)\epsilon$$

occurs.

As there is no component of the field parallel to v_0, the energy dispersion results only from the deviation Δv_z of the velocity component $v_z=dz/dt$, and thus produces a dispersion in the direction of the velocity. The calculations are straightforward and give for the directional dispersion:

$$\Delta\varphi_q\sim(e^3/m^3v_0)(k\hbar/\omega)^{\frac{1}{2}}E_0^{\frac{1}{2}}.$$

Consequently the ratio of the quantum dispersion in angle to the classical dispersion is

$$\Delta\varphi_q/\Delta\varphi_c\sim(k\hbar m\omega^3)^{\frac{1}{2}}(2\pi^2e)^{-1}E_0^{-1}\epsilon^{-1},$$

where k, an integer, is the number of periods spent by the electrons in the rf field. Taking for example 10-kev electrons, $\omega=2\pi\times10^{10}$ cps, $k=10$, and $\epsilon=10^{-3}$, we obtain

$$\Delta\varphi_q\sim10^{-3}E_0^{\frac{1}{2}}\text{ rad},\quad\text{and}\quad\Delta\varphi_q/\Delta\varphi_c\sim10^4E_0^{-\frac{1}{2}},$$

where E_0 is measured in volts/meter.

It is obviously desirable to have a very well focused beam and a long drift space when carrying out the experiment. However, with the aid of a simple collector system the expected effect can be studied quantitatively without making extremely high demands on focusing.[5] The current density in an electron beam produced by a conventional electron gun as a function of the distance from the beam axis is given by the Gaussian distribution law. Within the limits of our approximate estimate, the directional distribution due to the quantum effect may be assumed to follow the same law. Thus the resulting distribution is also of the same kind, and the ratio of the beam currents outside a circle of radius R with and without the rf field in the cavity is

$$I/I_0=1+2(R+r_0)^2(L/r_0)\Delta\varphi,$$

where L is the length of the drift space and r_0 is the distance from the beam axis to the point where the current density reduces to $1/e$ times its value on the axis.

As the directional spread $\Delta\varphi$ is proportional to E_0 in the classical case, and it is proportional to $E_0^{\frac{1}{2}}$ if the quantum nature of the interaction is dominant, the two effects can be distinguished without ambiguity if I/I_0 is plotted *versus* E_0.

Another method, followed by Senitzky[4,6] and others, consists in the application of quantum electrodynamics to the oscillations maintained in a resonant cavity. These considerations do not yield any dispersion of the kind given by Eq. (1). If the electron spends an integral number of periods in the cavity, the energy gain computed classically and quantum electrodynamically is zero simultaneously.[4] This means that the elementary emission and absorption acts should not be considered as independent processes. The velocity distribution obtained by Senitzky in the first approximation of the perturbation theory [Eq. (31) of reference 6] contains as a result of the quantization of the field only a vacuum fluctuation term, the other terms resulting from the wave mechanical treatment of the electrons.[7] It may be noted that the equations can be solved exactly for certain modes of the cavity (the electric field must be independent of the coordinate measured along the path of the electron). The results are in agreement with the first approximation obtained by Senitzky: the dispersion of the energy absorbed by the electrons during their transit time is independent of the field intensity and the number of periods spent within the cavity, if the effects characteristic of the wave-mechanical treatment of the electrons (spreading out of the wave packet) are neglected.

An experimental check of the velocity distribution given by Senitzky seems also to be more promising if a direction spread caused by a transverse rf field is sought. By following his method of calculation for a transverse rf field, the velocity spread is found to be

$$\begin{aligned}\langle v_z^2\rangle-\langle v_z\rangle^2&=\langle\Delta v_z^2\rangle-\langle\Delta v_z\rangle^2\\&=(\hbar^2/2m^2b^2)(e^2E_0^2u_0^2/v_0^2\omega_0)A^2\\&\quad+\tfrac{1}{2}b^2(e^2E_0^2u_0^2/mv_0^2)B^2,\end{aligned}\qquad(2)$$

instead of Eq. (31) of reference 6, if one neglects the field fluctuation term. If the number of periods spent by the electrons in the rf field is odd, then $B=0$. Thus by decreasing the quantity b, the velocity spread can be increased to an appreciable value. In spite of this favorable situation the experimental results could not be interpreted without ambiguity. With $\Delta r=b/\sqrt{2}$ and $\Delta v=\hbar/\sqrt{2}mb$, we get from Eq. (2) an expression for the resulting velocity spread which depends on Δr and Δv, the initial spread of coordinates and of velocities respectively, but does not contain Planck's constant. This means that the spreads of coordinates and velocities due to technical reasons (which are obviously greater than those of quantum mechanical origin) contribute to the directional spread in the same manner, and consequently the two kinds of effects cannot be distinguished.

Experiments based upon the above considerations are in progress to study in detail the validity of the relation (1), which is the consequence of the assumption of the independent character of the elementary processes. Our results will be published in the Acta Physica Hungarica at a later date.

We are indebted to Mr. F. Károlyházi for valuable discussions.

[4] I. R. Senitzky, Phys. Rev. **95**, 904 (1954).

[7] It is curious that according to this result the zero-point fluctuation of the field can be increased by decreasing the volume of the cavity.

Variational Calculation of Electron Scattering by a Static Potential*†

LYMAN MOWER‡
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received March 24, 1955)

The Schwinger variational method, for the approximate determination of scattering amplitude, is tested for accuracy in the case of the elastic scattering of electrons from the static potential $V(\mathbf{r})=-e^2V_0e^{-\lambda r}/r$, by using eight different forms of trial wave functions. The results are compared by checking the closeness of fit of the associated scattering amplitude with an exact solution to the problem. In the course of the calculation a number of expressions, of use in more complicated problems, were obtained and are here recorded. The parameter values used in the test were $a_0\lambda=8/3$, $k^2=(0.72\lambda)^2$, $V_0=7.8$, where a_0 is the first Bohr orbit radius for hydrogen.

1. INTRODUCTION

IN the application of variational methods to scattering problems the algebraic and numerical computations have been directed mainly toward the determination of either the first few phase shifts or the total scattering cross section.[1] However, for a given trial function the determination of the scattering amplitude imposes a stricter test on the reliability of the trial wave function than does the determination of the first few phase shifts or the total scattering amplitude.

We apply here the Schwinger variational method for the approximate determination of the scattering amplitude for the elastic scattering of electrons from a static potential.[2] Several different trial wave functions are compared for possible application to similar problems; the criterion used for "best form" consists in the closeness of fit of the associated scattering amplitude with a numerical solution to the problem.

For reference we include a brief description of the Schwinger variational method as applied to scattering problems.

2. PRELIMINARY REMARKS

The scattering of an electron by a potential $V(\mathbf{r})$ is described by a solution to the integral equation[3]

$$\psi_i(\mathbf{r})=e^{i\mathbf{k}_i\cdot\mathbf{r}}-\frac{2m}{\hbar^2}\int\frac{G(\mathbf{r},\mathbf{r}')}{4\pi}V(\mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}', \quad (1)$$

where

$$G(\mathbf{r},\mathbf{r}')=e^{ikR}/R, \quad R=|\mathbf{r}-\mathbf{r}'|, \quad (2)$$

is the free space Green function for the Helmholtz equation, $E=\hbar^2k^2/2m$ is the energy of the incident electron, m is its mass, and $\hbar$ is Plank's constant divided by 2π. The vector $\mathbf{k}_i=k\mathbf{n}_i$, where the unit vector $\mathbf{n}_i$ specifies the direction of incidence; the vector $\mathbf{r}=r\mathbf{n}$ is the radius vector which specifies the position of the electron. Substituting into (1) the asymptotic form of the Green function,[3]

$$G(\mathbf{r},\mathbf{r}')\rightarrow\exp(ikr-ik\mathbf{n}\cdot\mathbf{r}')/r, \quad r\gg r', \quad (3)$$

we obtain

$$\psi_i(\mathbf{r})\rightarrow e^{i\mathbf{k}_i\cdot\mathbf{r}}+f(\mathbf{n},\mathbf{n}_i)e^{ikr}/r, \quad r\gg r', \quad (4)$$

where

$$f(\mathbf{n},\mathbf{n}_i)=-\frac{m}{2\pi\hbar^2}\int\exp(-ik\mathbf{n}\cdot\mathbf{r}')V(\mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}' \quad (5)$$

is the scattering amplitude and is the quantity which we wish to characterize by a variational method. The quantity $|f(\mathbf{n},\mathbf{n}_i)|^2$ represents the density of the current scattered by the potential $V(\mathbf{r})$ in the direction $\mathbf{n}$ for a stream of particles of unit current density incident in the direction $\mathbf{n}_i$.

Using the integral equation (1) and the definition of the scattering amplitude (5), we may rewrite the expression for the scattering amplitude as[4]

$$[f(\mathbf{n}_s,\mathbf{n}_i)]=-\frac{1}{4\pi}\Big\{\int e^{-i\mathbf{k}_s\cdot\mathbf{r}'}U(\mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}'$$
$$+\int e^{i\mathbf{k}_i\cdot\mathbf{r}}U(\mathbf{r})\psi_{-s}(\mathbf{r})d\mathbf{r}-\int\psi_{-s}(\mathbf{r})U(\mathbf{r})\psi_i(\mathbf{r})d\mathbf{r}$$
$$-\frac{1}{4\pi}\int\int\psi_{-s}(\mathbf{r})U(\mathbf{r})G(\mathbf{r},\mathbf{r}')U(\mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}d\mathbf{r}'\Big\}, \quad (6)$$

where $U(\mathbf{r})=2mV(\mathbf{r})/\hbar^2$, and where the wave function

* This work was supported in part by the Office of Naval Research.

† This work is based in part on material submitted in partial fulfillment for the Ph.D. degree at Massachusetts Institute of Technology.

‡ Now at the Electronic Defense Laboratory, Box 205, Mountain View, California.

[1] L. Hulthén, Kgl. Fysiograf. Sallskap. Lund. Förh. **14**, 1 (1944); W. Kohn, Phys. Rev. **74**, 1763 (1948); S. S. Huang, Phys. Rev. **76**, 477 (1949); H. S. W. Massey and B. L. Moiseiwitsch, Proc. Roy. Soc. (London) **A205**, 483 (1951); G. A. Erskine and H. S. W. Massey, Proc. Roy. Soc. (London) **A212**, 521 (1952); S. Altschuler, Phys. Rev. **89**, 1278 (1953); H. Boyet and S. Borowitz, Phys. Rev. **93**, 1225 (1954).

[2] B. A. Lippman and J. Schwinger, Phys. Rev. **79**, 469 (1950); T. Kahan and G. Rideau, J. phys. radium **13**, 326 (1952); E. Gerjuoy and D. S. Saxon, Phys. Rev. **85**, 939 (1952) have used a form like the one listed as ϕ_{71} in Eq. (9).

[3] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Clarendon Press, Oxford, 1949), second edition, p. 116.

[4] See reference 2, and P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 804.

$\psi_{-s}(\mathbf{r})$ represents a solution to (1) for a plane wave incident in the direction $(-\mathbf{n}_s)$, where $\mathbf{k}_s=k\mathbf{n}_s$. We note that $[f(\mathbf{n}_s,\mathbf{n}_i)]$ is an exact expression for the scattering amplitude $f(\mathbf{n}_s,\mathbf{n}_i)$.

For the approximate determination of the scattering amplitude, the procedure followed is to substitute into the functional $[f(\mathbf{n}_s,\mathbf{n}_i)]$ a set of functions ϕ which satisfy or nearly satisfy the integral equation (1). These functions are termed trial functions. Then those trial functions which differ from the exact value of ψ to the first order in $\delta\psi$ yield values of the functional $[f(\mathbf{n}_s,\mathbf{n}_i)]$ which differ from the exact value of $[f(\mathbf{n}_s,\mathbf{n}_i)]$, i.e., the scattering amplitude, by terms proportional to $(\delta\psi)^2$. Hence, the difference between the scattering amplitude and the above value of the functional is small to the second order in $(\delta\psi)$.

By introducing trial functions of the form

$$\phi(\mathbf{r})=\sum_{n=0}^{N} C_n\chi_n(\mathbf{r}) \tag{7}$$

into (6), where the C_n are unknown constants and the $\chi_n(\mathbf{r})$ are known functions of $\mathbf{r}$, the functional $[f(\mathbf{n}_s,\mathbf{n}_i)]$ becomes a function of the C_n. Thus the necessary conditition that $[f(\mathbf{n}_s,\mathbf{n}_i)]$ be stationary with respect to small variations of $\phi_i(\mathbf{r})$ and $\phi_{-s}(\mathbf{r})$ about $\psi_i(\mathbf{r})$ and $\psi_{-s}(\mathbf{r})$ is that

$$\partial[f(\mathbf{n}_s,\mathbf{n}_i)]/\partial C_n=0, \quad n=0, 1, \cdots N. \tag{8}$$

The foregoing equations consist of a set of $N+1$ simultaneous, linear equations sufficient to determine uniquely the set of $N+1$ unknowns. By assuming a trial function of the form (7), the resultant expression for the scattering amplitude $[f(\mathbf{n}_s,\mathbf{n}_i)]$ is independent of the normalization of the trial wave function $\phi(\mathbf{r})$.

Clearly the method is only as good as the choice of the trial function. If we know the exact function, then we obtain the extremum value of $[f(\mathbf{n}_s,\mathbf{n}_i)]$, i.e., the scattering amplitude, exactly. However, if we make a bad choice in the trial function, then the error in the functional will be correspondingly augmented.

We note that inasmuch as the Green function (2) is not definite, the extremum point of the function $[f(\mathbf{n}_s,\mathbf{n}_i)]$ may be either a maximum, a minimum, or a saddle point.[5]

3. APPLICATION OF THE VARIATIONAL METHOD

For the trial wave functions

$$\begin{aligned}
&\phi_{i1}(\mathbf{r})=C_0e^{i\mathbf{k}_i\cdot\mathbf{r}},\\
&\phi_{i2}(\mathbf{r})=C_0e^{i\mathbf{k}_i\cdot\mathbf{r}}+C_1e^{i\mathbf{k}_i\cdot\mathbf{r}-\lambda r},\\
&\phi_{i3}(\mathbf{r})=C_0e^{i\mathbf{k}_i\cdot\mathbf{r}}+C_1e^{-\lambda r},\\
&\phi_{i4'}(\mathbf{r})=C_0e^{i\mathbf{k}_i\cdot\mathbf{r}}+C_1e^{-\lambda r}+iC_2\lambda e^{-\lambda r/2}(\mathbf{n}_i\cdot\mathbf{n}),\\
&\phi_{i5}(\mathbf{r})=C_0e^{i\mathbf{k}_i\cdot\mathbf{r}}+C_1j_0(kr),\\
&\phi_{i6}(\mathbf{r})=C_0e^{i\mathbf{k}_i\cdot\mathbf{r}}+C_1j_0(kr)+iC_2j_1(kr)(\mathbf{n}_i\cdot\mathbf{n}),\\
&\phi_{i7}(\mathbf{r})=C_0e^{i\mathbf{k}_i\cdot\mathbf{r}}+C_1e^{-i\mathbf{k}_i\cdot\mathbf{r}},\\
&\phi_{i8}(\mathbf{r})=C_0e^{i\mathbf{k}_i\cdot\mathbf{r}}+C_1e^{-i\mathbf{k}_i\cdot\mathbf{r}}+C_2j_0(kr),\\
&\phi_{i9'}(\mathbf{r})\to\phi_{i8}(\mathbf{r}) \text{ [see paragraph below Eq. (18)]},
\end{aligned} \tag{9}$$

where $j_n(x)$ is the nth order spherical Bessel function, and for the potential

$$V(\mathbf{r})=-e^2V_0e^{-\lambda r}/r, \tag{10}$$

approximate expressions for the scattering amplitude were calculated using the stationary form (6). We note that the trial wave functions $\phi_{-s}(\mathbf{r})$ are obtained from the $\phi_s(\mathbf{r})$ by replacing the vector $\mathbf{k}_i$ by $(-\mathbf{k}_s)$; for simplification, we have introduced the notation

$$\begin{aligned}
&[f(\theta)]=[f(\mathbf{n}_s,\mathbf{n}_i)],\\
&\theta=\arccos(\mathbf{n}_i\cdot\mathbf{n}_s),\\
&x=2k/\lambda,\\
&c=\sin(\theta/2).
\end{aligned} \tag{11}$$

As an illustration of the procedure we calculate $[f(\theta)]$ for the trial wave function ϕ_7 [see (9)]; thus, upon performing the indicated integrations, we obtain

$$\begin{aligned}
[f_7(\theta)]=f_B(\theta)\{&2C_0+2C_1\alpha-(C_0^2+C_1^2)[1-T_0(\theta)]\\
&-2C_0C_1\alpha[1-T_0(\pi-\theta)]\},
\end{aligned} \tag{12}$$

where[5]

$$\begin{aligned}
f_B(\theta)&=-\frac{1}{4\pi}\int e^{i(\mathbf{k}_i-\mathbf{k}_s)\cdot\mathbf{r}}U(\mathbf{r})d\mathbf{r}\\
&=(U_0/\lambda)(x^2c^2+1)^{-1}
\end{aligned} \tag{13}$$

is the first Born approximation for the potential considered, and

$$\begin{aligned}
&U(\mathbf{r})=-(U_0\lambda/r)e^{-\lambda r}, \quad U_0=2V_0/a_0\lambda\\
&\alpha=f_B(\pi-\theta)/f_B(\theta)=(x^2c^2+1)/(x^2-x^2c^2+1),\\
&T_0(\theta)=\frac{U_0^2\lambda^2}{f_B(\theta)}\left\{\frac{1}{2\pi^2}\int\frac{d\boldsymbol{\eta}}{(\eta^2-k^2)[(\boldsymbol{\eta}-\mathbf{k}_i)^2+\lambda^2][(\boldsymbol{\eta}-\mathbf{k}_s)^2+\lambda^2]}\right\}\\
&\quad=U_0(1+x^2c^2)\left\{\frac{2}{xcD}\left[\arctan\left(\frac{xc}{D}\right)+\frac{i}{2}\ln\frac{D+x^2c}{D-x^2c}\right]\right\},\\
&D^2=4+4x^2+x^4c^2,
\end{aligned} \tag{14}$$

[5] See reference 4, p. 1149.

where $a_0=\hbar^2/mc^2$ is the first Bohr radius for hydrogen. Upon applying (8) to (12) and solving the resultant equations for C_0 and C_1, we find

$$\begin{aligned} C_0&=\{[1-T_0(\theta)]-\alpha^2[1-T_0(\pi-\theta)]\}/\\ &\qquad\{[1-T_0(\theta)]^2-\alpha^2[1-T_0(\pi-\theta)]^2\},\\ C_1&=\alpha\{[1-T_0(\theta)]-[1-T_0(\pi-\theta)]\}/\\ &\qquad\{[1-T_0(\theta)]^2-\alpha^2[1-T_0(\pi-\theta)]^2\}. \end{aligned} \tag{15}$$

Hence, upon substitution of these expressions into (12), we find that the scattering amplitude is given by

$$[f_7(\theta)]=f_B(\theta)\frac{(\alpha^2+1)[1-T_0(\theta)]-2\alpha^2[1-T_0(\pi-\theta)]}{[1-T_0(\theta)]^2-\alpha^2[1-T_0(\pi-\theta)]^2}, \tag{16}$$

which at $\theta=\pi/2$ reduces to

$$[f_7(\pi/2)]=f_B(\pi/2)\frac{1}{1-T_0(\pi/2)}. \tag{17}$$

The following stationary values of the scattering amplitude $[f(\theta)]$ for the various trial wave functions listed in (9) were obtained:

$$\begin{aligned}
&\phi_{t1}\colon\ f_B(\theta)\left\{\frac{1}{1-T_0(\theta)}\right\},\\
&\phi_{t2}\colon\ f_B(\theta)\left\{\frac{d_0^2[1-T_0(\theta)]-2d_0(d_0-T_1)+(d_1-T_2)}{[1-T_0(\theta)](d_1-T_2)-(d_0-T_1)^2}\right\},\\
&\phi_{t3}\colon\ f_B(\theta)\left\{\frac{d_2[1-T_0(\theta)]+(d_4-T_4)-2d_2(d_2-T_3)}{[1-T_0(\theta)](a_4-T_4)-(a_2-T_3)^2}\right\},\\
&\phi_{t4'}\colon\ f_B(\theta)\left\{\frac{d_2[1-T_0(\theta)]+(d_4-T_4)-2d_2(d_2-T_3)}{[1-T_0(\theta)](d_4-T_4)-(d_2-T_3)^2}\right\}+f_B(\theta)\frac{\cos\theta}{d_6-T_6}\left\{d_2-(d_2-T_3)\frac{[(d_4-T_4)-d_2(d_2-T_3)]}{[1-T_0(\theta)](d_4-T_4)-(d_2-T_3)^2}\right\}^2,\\
&\phi_{t5}\colon\ f_B(\theta)\left\{\frac{d_6^2}{d_6-T_7}+\frac{(1-d_6)^2}{[1-T_0(\theta)]-(d_6-T_7)}\right\},\\
&\phi_{t6}\colon\ f_B(\theta)\left\{\frac{d_6^2}{d_6-T_7}+\frac{3d_7^2\cos\theta}{d_7-T_8}+\frac{(1-d_6-3d_7\cos\theta)^2}{[1-T_0(\theta)]-(d_6-T_7)-3(d_7-T_8)\cos\theta}\right\},\\
&\phi_{t7}\colon\ f_B(\theta)\left\{\frac{(\alpha^2+1)[1-T_0(\theta)]-2\alpha^2[1-T_0(\pi-\theta)]}{[1-T_0(\theta)]^2-\alpha^2[1-T_0(\pi-\theta)]^2}\right\},\\
&\phi_{t8}\colon\ f_B(\theta)\left\{\frac{2(\alpha-d_6)(1-d_6)}{[1-T_0(\theta)]+\alpha[1-T_0(\pi-\theta)]-2(d_6-T_7)}\right.\\
&\qquad\left.+\frac{(1-\alpha^2)\{(d_6-T_7)-[1-T_0(\theta)]\}}{\{\alpha[1-T_0(\pi-\theta)]-[1-T_0(\theta)]\}\{[1-T_0(\theta)]+\alpha[1-T_0(\pi-\theta)]-2(d_6-T_7)\}}+\frac{d_6^2}{d_6-T_7}\right\},\\
&\phi_{t9'}\colon\ f_B(\theta)\left\{\frac{(\alpha^2+1)[1-T_0(\theta)]-2\alpha^2[1-T_0(\pi-\theta)]}{[1-T_0(\theta)]^2-\alpha^2[1-T_0(\pi-\theta)]^2}\right\}+f_B(\theta)(d_8-T_7)\left\{\frac{d_8}{d_8-T_7}-\frac{1+\alpha}{1-T_0(\theta)+\alpha[1-T_0(\pi-\theta)]}\right\}^2.
\end{aligned} \tag{18}$$

The definitions of the symbols used are included in the Appendix. For two of the trial wave functions ($\phi_{t4'}$ and $\phi_{t9'}$) we have determined the value of the constant C_2 by (8) while taking for the values of C_0 and C_1 those of the previous wave functions (ϕ_{t3} and ϕ_{t8}). To denote this procedure, we insert a prime next to the number of the trial wave function.

Numerical calculations were performed for the differential scattering cross section $|[f(\theta)]|^2$ for each of the algebraic expressions listed in (18); these were then compared with a numerical solution obtained by phase shift analysis. The values of the parameters U_0, λ, and x were taken as

$$\begin{aligned} U_0&=3V_0/4=5.85,\\ \lambda&=8/3a_0,\\ x&=1.44, \end{aligned} \tag{19}$$

where a_0 designates the first Bohr radius for the hydrogen atom. (These values were chosen for future

reference for the elastic scattering of electrons from neon.) By numerical solution of the appropriate radial wave equation,[6] the following phase shifts were obtained:

δ_0	δ_1	δ_2	δ_3
188.0°	50.9°	6.3°	1.2°

(The phase shift δ_3 was calculated by a first Born approximation.[7])

Inasmuch as $|f(\theta)|^2$ was expected to be a relatively smooth function of θ, numerical expressions for each of the functions listed in (18) were calculated for only three values of θ: 0, $\frac{1}{2}\pi$, and π. The following numerical values were obtained for the differential scattering cross section, $|[f(\theta)]|^2/a_0^2$, for the various trial wave functions listed:

θ	Num. soln.	ϕ_1	ϕ_2	ϕ_3	$\phi_{4'}$	ϕ_5	ϕ_6	ϕ_7	ϕ_8	$\phi_{9'}$
0	2.41	2.58	0.92	5.31	4.49	2.12	2.14	2.39	2.45	2.42
$\frac{1}{2}\pi$	0.005	0.17	0.03	0.13	0.13	0.28	0.28	0.17	0.28	0.17
π	1.05	0.04	0.15	0.05	0.28	0.29	0.33	1.01	0.47	0.99

From and inspection of the above table we see that the values of $|[f(\theta)]|^2$ at $x=1.44$ associated with ϕ_7 and $\phi_{9'}$ are comparable and, for the trial wave functions chosen, have the "best form." We note that ϕ_7 is the simpler of these two trial wave functions.

ACKNOWLEDGMENTS

The writer is grateful to Professor Philip M. Morse for suggesting the problem and for his understanding and helpful guidance.

APPENDIX. TABULATION OF VARIOUS INTEGRALS

The definitions of the functions d_n, α, and $f_B(\theta)$ which arise from the first three integrals in (6) are as follows:

$$d_0=-\frac{1}{4\pi f_B(\theta)}\int e^{i(\mathbf{k}_i-\mathbf{k}_s)\cdot\mathbf{r}}e^{-\lambda r}U(\mathbf{r})d\mathbf{r}=\frac{x^2c^2+1}{x^2c^2+4},$$

$$d_1=-\frac{1}{4\pi f_B(\theta)}\int e^{i(\mathbf{k}_i-\mathbf{k}_s)\cdot\mathbf{r}}e^{-2\lambda r}U(\mathbf{r})d\mathbf{r}=\frac{x^2c^2+1}{x^2c^2+9},$$

$$d_2=-\frac{1}{4\pi f_B(\theta)}\int e^{i\mathbf{k}_i\cdot\mathbf{r}}e^{-\lambda r}U(\mathbf{r})d\mathbf{r}=\frac{4(x^2c^2+1)}{x^2+16},$$

$$d_3=-\frac{i\lambda}{4\pi f_B(\theta)\cos\theta}\int e^{i\mathbf{k}_i\cdot\mathbf{r}}e^{-\lambda r/2}U(\mathbf{r})(\mathbf{n}_s\cdot\mathbf{r})d\mathbf{r}=\frac{16x(x^2c^2+1)}{(9+x^2)^2},$$

$$d_4=-\frac{1}{4\pi f_B(\theta)}\int e^{-2\lambda r}U(\mathbf{r})d\mathbf{r}=\frac{x^2c^2+1}{9},$$

$$d_5=-\frac{\lambda^2}{12\pi f_B(\theta)}\int e^{-\lambda r}U(\mathbf{r})r^2d\mathbf{r}=\frac{x^2c^2+1}{8},$$

$$d_6=-\frac{1}{4\pi f_B(\theta)}\int e^{i\mathbf{k}_i\cdot\mathbf{r}}U(\mathbf{r})j_0(kr)d\mathbf{r}=\frac{x^2c^2+1}{x^2}\ln(1+x^2),$$

$$d_7=-\frac{i}{4\pi f_B(\theta)\cos\theta}\int e^{i\mathbf{k}_i\cdot\mathbf{r}}U(\mathbf{r})j_1(kr)(\mathbf{n}_s\cdot\mathbf{r})d\mathbf{r}=\frac{2(x^2c^2+1)}{x^2}Q_1\left(1+\frac{2}{x^2}\right),$$

$$f_B(\theta)=-\frac{1}{4\pi}\int U(\mathbf{r})e^{i(\mathbf{k}_i-\mathbf{k}_s)\cdot\mathbf{r}}d\mathbf{r}=\frac{U_0}{\lambda}\frac{1}{x^2c^2+1},$$

$$\alpha=\frac{f_B(\pi-\theta)}{f_B(\theta)}=\frac{x^2c^2+1}{x^2(1-c^2)+1},$$

$$U(\mathbf{r})=-U_0(\lambda/r)e^{-\lambda r},\quad x=2k/\lambda,\quad c=\sin(\theta/2),$$

$$Q_1(z)=\tfrac{1}{2}z\ln\left(\frac{z+1}{z-1}\right)-1,\quad z>1.$$

The evaluation of the double integral which occurs in (6) is simplified by a transformation to momentum space; thus, the Green function is replaced by its Fourier transform

$$G(\mathbf{r},\mathbf{r}')=\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}=\frac{4\pi}{(2\pi)^3}\int\frac{d\boldsymbol{\eta}\exp[i\boldsymbol{\eta}\cdot(\mathbf{r}-\mathbf{r}')]}{\eta^2-k^2},\quad\text{(A.1)}$$

where the path of integration is defined to be over the pole at $\eta=-k$ and under that at $+k$. Hence, double integrals of the form

$$\int\int J_1(\mathbf{r},-\mathbf{k}_s)G(\mathbf{r},\mathbf{r}')J_2(\mathbf{r}',\mathbf{k}_i)d\mathbf{r}d\mathbf{r}'\quad\text{(A.2)}$$

are reduced to integrals of the form

$$4\pi\int\frac{d\boldsymbol{\eta}}{\eta^2-k^2}K_1(\boldsymbol{\eta},-\mathbf{k}_s)K_2(-\boldsymbol{\eta},\mathbf{k}_i),\qquad K(\boldsymbol{\eta},\mathbf{k})=(2\pi)^{-1}\int J(\mathbf{r},\mathbf{k})e^{i\boldsymbol{\eta}\cdot\mathbf{r}}d\mathbf{r}.\quad\text{(A.3)}$$

In several of the double integrals the resultant algebraic expressions are simplified if, prior to integration, a coordinate transformation is carried out[8] or Feynman's method[9] of grouping factors is used.

Thus, the definitions of the functions $T_n(\theta)$, which

[6] See reference 3, p. 129.
[7] See reference 3, pp. 28 and 119.
[8] G. Källén, Arkiv Fysik **2**, 33 (1950).
[9] R. H. Dalitz, Proc. Roy. Soc. (London) **A206**, 509 (1950).

arise from the double integral in (6) are as follows:

$$T_0(\theta,x)=\frac{U_0^2\lambda^2}{2\pi^2 f_B(\theta)}\int\frac{d\boldsymbol{\eta}}{(\eta^2-k^2)[(\boldsymbol{\eta}-\mathbf{k}_i)^2+\lambda^2][(\boldsymbol{\eta}-\mathbf{k}_s)^2+\lambda^2]}=\left(\frac{U_0^2}{\lambda f_B(\theta)}\right)\frac{2}{xcD}\left[\tan^{-1}\frac{xc}{D}+\frac{i}{2}\ln\frac{D+x^2c}{D-x^2c}\right],$$

$$T_1(\theta)=\frac{U_0^2\lambda^2}{2\pi^2 f_B(\theta)}\int\frac{d\boldsymbol{\eta}}{(\eta^2-k^2)[(\boldsymbol{\eta}-\mathbf{k}_i)^2+4\lambda^2][(\boldsymbol{\eta}-\mathbf{k}_s)^2+\lambda^2]}=\left(\frac{U_0^2}{\lambda f_B(\theta)}\right)\frac{1}{xb_1}\left\{\left[\tan^{-1}\frac{xb_1}{3-x^2c^2}-\tan^{-1}\frac{xb_1}{2(3+x^2c^2)}\right]\right.$$
$$\left.+\frac{i}{2}\ln\left[\frac{16b_2^2+x^2(b_3+b_1)^2}{16b_2^2+x^2(b_3-b_1)^2}\right]\cdot\left[\frac{4b_2^2+x^2(b_4+b_1)^2}{4b_2^2+x^2(b_4-b_1)^2}\right]\right\},$$

$$T_2(\theta)=\frac{x^2c^2+1}{2(x^2c^2+4)}T_0(\theta,x/2),$$

$$T_3(\theta)=\frac{U_0^2\lambda^2}{2\pi^2 f_B(\theta)}\int\frac{d\boldsymbol{\eta}}{(\eta^2-k^2)[(\boldsymbol{\eta}-\mathbf{k}_i)^2+\lambda^2][\eta^2+4\lambda^2]}=\left(\frac{U_0^2}{\lambda f_B(\theta)}\right)\frac{4}{x(x^2+16)}\left[\tan^{-1}x-2\tan^{-1}\frac{x}{6}+\frac{i}{2}\ln(1+x^2)\right],$$

$$T_4(\theta)=\frac{U_0^2\lambda^2}{2\pi^2 f_B(\theta)}\int\frac{d\boldsymbol{\eta}}{(\eta^2-k^2)(\eta^2+4\lambda^2)^2}=\left(\frac{U_0^2}{\lambda f_B(\theta)}\right)\left[\frac{16-x^2+i8x}{(16+x^2)^2}\right],$$

$$T_5(\theta)=\frac{U_0^2\lambda^3}{4\pi^2 f_B(\theta)\cos\vartheta}\int\frac{\mathbf{n}_s\cdot\boldsymbol{\eta}\,d\boldsymbol{\eta}}{(\eta^2-k^2)(\eta^2+9\lambda^2/4)^2[(\mathbf{k}_i-\boldsymbol{\eta})^2+\lambda^2]}$$
$$=-\left(\frac{U_0^2}{\lambda f_B(\theta)}\right)\frac{4}{x(x^2+9)^2}\left\{i\left[x-\frac{2+x^2}{2x}\ln(1+x^2)\right]+\frac{2(15-x^2)}{25+x^2}+\frac{(2+x^2)}{x}[2\tan^{-1}(x/5)-\tan^{-1}x]\right\},$$

$$T_6(\theta)=\frac{U_0^2\lambda^4}{8\pi^2 f_B(\theta)\cos\vartheta}\int\frac{(\mathbf{n}_s\cdot\boldsymbol{\eta})(\mathbf{n}_i\cdot\boldsymbol{\eta})d\boldsymbol{\eta}}{(\eta^2-k^2)(\eta^2+9\lambda^2/4)^4}=\left(\frac{U_0^2}{\lambda f_B(\theta)}\right)\frac{1}{3(9+x^2)^4}\left[i8x^3-\frac{(x^6+81x^4-729x^2-729)}{54}\right],$$

$$T_7(\theta)=\frac{U_0^2}{8\pi^2x^2f_B(\theta)}\int\frac{d\boldsymbol{\eta}}{\eta^2(\eta^2-k^2)}\left\{\ln\frac{(\eta+k)^2+\lambda^2}{(\eta-k)^2+\lambda^2}\right\}^2=\left(\frac{U_0^2}{\lambda f_B(\theta)}\right)\frac{2}{x^3}\left\{\frac{i}{4}[\ln(1+x^2)]^2\right.$$
$$\left.+(\tfrac{1}{2}\pi-\tan^{-1}x)\ln(1+x^2)+\sum_{n=1}^{\infty}\frac{\sin(2n\tan^{-1}x)}{n^2(1+x^2)^n}-\sum_{n=1}^{\infty}\frac{\sin(2n\tan^{-1}x)}{n^2}\right\},$$

$$T_8(\theta)=\frac{U_0^2}{2\pi^2x^2f_B(\theta)}\int\frac{d\boldsymbol{\eta}}{\eta^2(\eta^2-k^2)}\left[Q_1\left(\frac{\lambda^2+k^2+\eta^2}{2k\eta}\right)\right]^2=\left(\frac{U_0^2}{\lambda f_B(\theta)}\right)\frac{2}{x^3}\left\{i\left[Q_1\left(\frac{2+x^2}{x^2}\right)\right]^2\right.$$
$$+\frac{4}{x^2}\ln\left(1+\frac{x^2}{4}\right)-2\left(\frac{x^2+2}{x^2}\right)\tan^{-1}\left(\frac{x}{2+x^2}\right)$$
$$\left.+\left(\frac{2+x^2}{x^2}\right)^2\left[\left(\frac{\pi}{2}-\tan^{-1}x\right)\ln(1+x^2)+\sum_{n=1}^{\infty}\frac{\sin(2n\tan^{-1}x)}{n^2(1+x^2)^n}-\sum_{n=1}^{\infty}\frac{\sin(2n\tan^{-1}x)}{n^2}\right]\right\},$$

$$c=\sin(\theta/2),\qquad b_1^2=(x^2c^2+5)^2+16c^2-16,$$
$$x=2k/\lambda,\qquad b_2=3-x^2c,$$
$$D^2=4+4x^2+x^4c^2,\qquad b_3=x^2c^2-8c-3,$$
$$Q_1(z)=\tfrac{1}{2}z\ln\left(\frac{z+1}{z-1}\right)-1,\qquad b_4=x^2c^2+2c+3.$$

Effect of Radiation from Solar Flares on the Ionosphere and the Earth's Magnetic Field

R. N. SEDRA AND I. B. HAZZAA
Faculty of Science, Cario University, Giza, Egypt
(Received February 11, 1955)

Variation in the earth's magnetic field and the corresponding variations in the ionic density of the F_1 layer due to solar flares are recorded. The value of the intensity of ultraviolet radiation before it enters the atmosphere is obtained. The constant relating the ionic density and the accompanying magnetic variation is also estimated.

I. INTRODUCTION

ULTRAVIOLET radiation from solar flares affects the ionic density of the E and F_1 layers of the ionosphere. It is also established that solar flares are usually accompanied by synchronous magnetic disturbances called "crochets."

In a previous communication[1] a relation between the intensity of ultraviolet radiation from solar flares and the accompanying magnetic disturbances was found.

Assuming that these "crochets" are intensifications of duirnal magnetic variations and that the solar ultraviolet flash produces an ion cloud, its magnetic effect on the earth could be calculated by replacing it by a bar magnet.[2]

The aim of this paper is to determine experimentally, from ionospheric and magnetic data, the value of the constant relating the ionic density and the magnetic disturbance.

THEORETICAL CONSIDERATION

Following the assumption of Maris and Hulburt[2] in taking the end S of the magnet to be 300 km thick and 1000 km wide, the distance of the ionized layer from the earth to be 200 km, and neglecting the north pole, the field F due to S is given by[1]

$$F=-(nkT/\pi H)\times 2.81, \tag{1}$$

where n is the ion density, k Boltzmann's constant, T the absolute temperature, and H the magnetic field.

Since T and H are constants for a certain height of the ionosphere, Eq. (1) could be put in the form:

$$n=-KF, \tag{2}$$

where K is a constant. By taking observations of the variation in the ionic density due to a solar flare and the corresponding magnetic disturbance, K could be evaluated experimentally.

APPARATUS

Simultaneous observations have been taken by magnetic and ionospheric recorders.

(a) A saturated core magnetometer is used for recording variations in the horizontal component of the earth's magnetic field.

Its circuit diagram is shown in Fig. 1. It consists of primary and secondary coils C_1 and C_2 provided with mumetal core. A sinusoidal voltage of frequency 1000 cycles/sec is applied to the primary coils. The second harmonic is filtered from the output of the secondary coils, amplified by an ac amplifier, then rectified and amplified again by a dc amplifier. The output is applied to the grid of a double triode.

The coils are placed parallel to the direction of the horizontal component of the earth's magnetic field and the output on the double triode is compensated by a potentiometer arrangement.

Any variation in the earth's magnetic field produces an output which is recorded by a milliammeter recorder connected between the anodes of the double triode.

(b) The automatic ionospheric recorder, built by Union Radio Company in England, consists of the following basic units.

(i) A pulse transmitter covering completely the frequency range 0.7 Mc/sec to 25 Mc/sec. The pulse width could be varied from 100 μsec to 330 μsec; and the peak output power is of the order 1.2 kw. The radio-frequency power is delivered into a rhombic aerial.

(ii) A conventional receiver covering the same range as the transmitter. It is provided with a servo-control system for keeping the transmitter and receiver in step as the transmitter is varied.

(iii) A crystal-controlled frequency generator providing suitable signals for frequency calibration and height marks, which are derived from a 3-kc/sec output

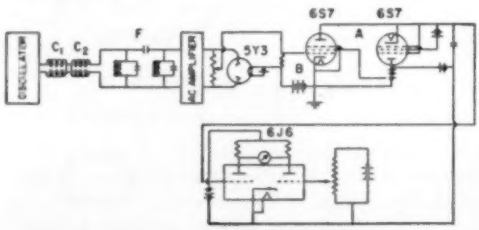


FIG. 1. Circuit diagram of the saturated core magnetometer showing connections of oscillator, coils, ac amplifier, rectifier, dc amplifier, compensator, and milliammeter recorder.

[1] R. N. Sedra, Phys. Rev. **83**, 190 (1951).
[2] H. B. Maris and E. O. Hulburt, Phys. Rev. **33**, 412 (1929).

giving representation of 50-km intervals along the time base of two cathode-ray tube displays (one for visual monitoring and the other for photographic recordings; associated with a camera using 70-mm film).

EXPERIMENTAL RESULTS

(1) Calculation of the Intensity of Solar Ultraviolet Radiation before it Enters the Earth's Atmosphere

The electron density, N, in the E and F_1 layers is given by the relation[3]

$$\frac{dN}{dt}=\frac{\beta I_0}{(kT/mg)\exp(1)}\cos\chi-\alpha N^2 \qquad (3)$$

$$=C\cos\chi-\alpha N^2,$$

where β is the number of ions produced by the absorption of a unit quantity of radiation, I_0 the intensity of solar ultraviolet radiation before it enters the atmosphere, m the mean molecular mass, χ the solar zenith angle, and α the effective recombination coefficient.

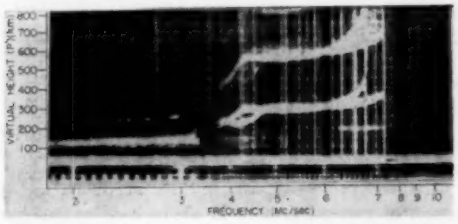


FIG. 2. Photograph obtained by the automatic ionospheric recorder showing echoes from E, F layers and ionic cloud detected at a height of 200 km and of mean ionic density 4.7×10^5 ions/cm^3.

From the $(P'-f)$ curves (P' being the virtual height and f the frequency) recorded by the ionospheric recorder, N is calculated from the relation $N=1.248\times10^4 f_c^2$, f_c being the critical frequency. After obtaining the values of N at intervals during the day, Eq. (3) is solved graphically and a value of C is obtained. By substituting the values of the constants, I_0 could be evaluated.

The experimental value of C for the E-layer ions was found to be 48 ions/cm^3 sec. Putting $\beta=1/(14\times1.6\times10^{-12})$, $m=4.8\times10^{-23}$ g, $g=900$ cm/sec^2, and $T=280°$ K at height 110 km, the value of $I_0=2.8\times10^{-3}$ erg cm^{-2} sec^{-1}.

(2) Temperature of the F_1 Layer from Ionospheric Data

From the $(P'-f)$ records obtained for the F_1 layer, the same procedure of graphical solution of Eq. (3) yields a value of C for the F_1 layer of 100 ions/cm^3 sec. By substituting the values of I_0, β, m, and g, the temperature of the F_1 layer is found to be 580°K.

[3] S. Chapman, Proc. Phys. Soc. (London) 43, 26 and 433 (1931).

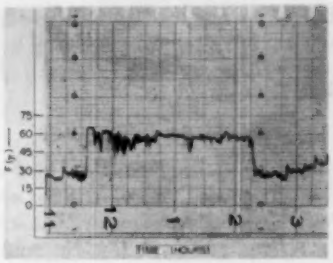


FIG. 3. A record obtained by the magnetic recorder observed simultaneously with the ionic cloud shown in Fig. 2. The mean variation in magnetic intensity amounts to 28×10^{-4} gauss.

(3) Relation between the Cloud Ionic Density and the Variation in the Earth's Magnetic Field

It has been observed in the $(P'-f)$ records of the ionospheric equipment, that abnormal increases in ion density occurred for short periods during the day as shown in Fig. 2. Simultaneously sudden variations in magnetic intensity were recorded by the magnetic recorder. This is illustrated in Fig. 3.

From the $(P'-f)$ records the values of n were calculated, and the corresponding changes in the earth's magnetic field F were measured from the magnetic records. Hence the value of the constant K in Eq. (2) was determined from the experimental data. Table I shows the results obtained. Mean value of $K=1.9\times10^9$ ions/cm^3/gauss.

From Eqs. (1) and (2)

$$K=(\pi H/kT)\times(1/2.81). \qquad (4)$$

The value of the magnetic field at 200 km is obtained from the known equation:

$$H=(2\pi mc/e)(f_x-f_0), \qquad (5)$$

where f_0 and f_x are the critical frequencies of the ordinary and extraordinary rays, respectively. If we substitute the values of f_0, f_x and the constants in Eq. (5), the value of H amounts to 0.21 gauss. Then, if we put the experimental values obtained for H, T, and

TABLE I. The values of ionic density and the corresponding changes in the horizontal component of the earth's magnetic field during the period from April, 1953, to October, 1954. The last column gives the value of K calculated from Eq. (2).

Date	Hour	n ions/cm^3	F gauss	K ion/cm^3 gauss
April 1, 1953	9	1.2×10^5	7×10^{-4}	1.7×10^9
April 2, 1953	11	1.4×10^5	8×10^{-4}	1.8×10^9
April 4, 1953	7	9.1×10^4	5×10^{-4}	1.8×10^9
April 14, 1953	14	2.0×10^5	10×10^{-4}	2.0×10^9
July 8, 1954	8	2.1×10^5	12×10^{-4}	1.9×10^9
August 14, 1954	7	1.1×10^5	6×10^{-4}	1.8×10^9
September 13, 1954	9	1.9×10^5	10×10^{-4}	1.9×10^9
October 16, 1954	11	3.9×10^5	20×10^{-4}	1.9×10^9
October 19, 1954	11	4.7×10^5	28×10^{-4}	1.7×10^9

K into Eq. (4), we find that the value of the constant 2.81, which depends on the dimensions of the effective magnet assumed by Maris and Hulburt, is modified to 4.4×10^{9}.

DISCUSSION OF RESULTS

The value of the normal intensity of solar ultraviolet radiation was calculated by Hulburt[4] and was 8×10^{-3} erg cm^{-2} sec^{-1}. Gledhill and Szendrei[5] obtained a mean value of 0.313 erg cm^{-2} sec^{-1}. Our experimental value amounts to 2.8×10^{-3} erg cm^{-2} sec^{-1} which is in closer agreement with the calculated value of Hulburt.

Using our value of I_0 to calculate the temperature of the F_1 layer at a height of 200 km, we obtain a value of 580°K which agrees with previous estimates. This proves that our value of I_0 is reasonable.

The numerical coefficient in Eq. (1) was found to have the value 4.4×10^{9} rather than the value 2.81 which was adopted by Maris and Hulburt. The discrepancy between the two values may be ascribed to the assumed dimensions of the effective magnet.

If we put our experimental value of the constant K in Eq. (2), the direct relation between the ionic density of the cloud and the accompanied magnetic variation is given by : $n=-1.9\times10^{9}F$.

[4] E. O. Hulburt, *Physics of the Earth*, edited by J. A. Fleming (McGraw-Hill Book Company, Inc., New York, 1949), Chap. X, p. 492.
[5] J. A. Gledhill, and M. E. Szendrei, Proc. Phys. Soc. (London) **B63**, 427 (1950).

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Growth of Current between Parallel Plates

P. M. DAVIDSON
University College of Swansea, Swansea, Wales
(Received April 22, 1955)

The author's exact solution of the mathematical problem of ionization current growth involving cathode secondary processes is here given in an alternative form which is more convenient for discussing the earliest stages of the growth.

FOR assessing the relative magnitude of the various regenerative processes which can occur in a gaseous discharge, considerable importance attaches to the theory of current growth. Of particular interest is the growth of current, in absence of space charge distortion of the field, between parallel plates maintained at a constant potential difference, and with a constant electron generation I_0 maintained at the cathode by external illumination, and with secondary electron generation due to the action of photons and positive ions on the cathode. The writer has given[1] the exact solution of this problem, and also an approximate solution. Bartholomeyczyk had given an approximate solution which neglected the maintained I_0, and was thus inapplicable to the present problem, except for tracing, very roughly, the current growth in case of high overvoltage. In the writer's approximate solution (which does take account of the maintained I_0) the electron current $I_-(0,t)$ at $x=0$ (the cathode) at time t is of the form

$$A-Be^{\lambda t}, \tag{1}$$

and the constant B is not specified exactly.

A recent paper by Bandel,[2] describing his experimental work on current growth and its interpretation, is of considerable interest, but his mathematical discussion of current growth in absence of space-charge distortion is inaccurate.[3] He considers the case in which, at time zero, there are no charged particles between the plates [the case called Part (i) in my exact solution]. He regards my exact solution as impracticably cumbersome; on the other hand he maintains that at times greater than d/v (or d/v_- if γ is zero[4]) my approximate solution, with a correct choice of the constant B is the exact solution of the problem. This contention is not correct, and the mathematical argument by which it is obtained is erroneous, for the following reason.

The integral equation (5) in Bandel's paper specifies the desired solution throughout all (positive) times. My approximate expression (1), with any value of B, satisfies the integral equation at t's greater than a specified value; but it does not follow that, with a correctly chosen B, it is the desired solution in that range of t. It can, in fact, be shown that the differential equations and boundary conditions [Eqs. (1) and (2) in Bandel's paper] require that if the expression (1) holds exactly after a finite time it must have held exactly at all (positive) times. Moreover $I_-(x,0)$ must have been, not zero as the problem requires, but

[1] Dutton, Haydon, and Jones, Brit. J. Appl. Phys. **4**, 170 (1953), Mathematical Appendix by P. M. Davidson.
[2] H. W. Bandel, Phys. Rev. **95**, 1117 (1954).
[3] For his purposes, the resulting error is not at all serious in his ultimate numerical applications, which are confined to the case in which positive ions do not generate electrons at the cathode.
[4] This notation is that of my previous paper (reference 1).

$e^{\alpha x}\{A-Be^{-\lambda x/v_-}\}$. Except in the case when γ is zero, $I_+(x,0)$ must also have been a nonzero distribution. Such facts are hardly surprising, for, as was pointed out in my paper, the equation for λ has, in addition to its obvious root which appears in (1) an infinite number of others, which appear in my exact solution. It would be surprising if the desired solution, in a certain range of time, could be constructed exactly by using only the one root.

In support of his contention, Bandel gives expressions which he says are the exact solution in the case of δ zero. It may readily be verified, however, that this is not so; there is a range of t (between d/v and $2d/v$) in which his expressions fail, and the alteration which they need in that range leads to alterations in all later ranges. This may be expressed explicitly by writing my exact solution in an alternative form. Although for most purposes it is, as we shall see, less convenient than the original form, it demonstrates the present point. For δ zero, $I_-(0,t)$ in the range nd/v to $(n+1)d/v$ is

$$I_0\sum_{m=0}^{n}\gamma^m e^{m\alpha d}\left\{(1+\gamma)^{-(m+1)}+\frac{(-\alpha v)^m}{m!}\times\left[\frac{d^m}{dp^m}\left\{\left(1-\frac{\alpha v}{p}\right)e^{p(t-md/v)}\right\}\right]_{p=p_0}\right\}, \quad (2)$$

where $p_0=\alpha v(1+\gamma)$ and 0! means unity. Thus when $t<d/v$, it is $I_0(1+\gamma e^{p_0 t})/(1+\gamma)$, and when $d/v<t<2d/v$, it is

$$\frac{I_0}{1+\gamma}\left\{\left(1+\frac{\gamma e^{\alpha d}}{1+\gamma}\right)+\gamma e^{p_0 t}\left(1-\frac{\{1+\gamma p_0(t-d/v)\}e^{-\gamma\alpha d}}{1+\gamma}\right)\right\}$$

and so on (extra terms being added in each successive range). Comparing this with Bandel's expression, we see that the latter is correct only up to $t=d/v$; as soon as it becomes identical with my approximate expression it ceases to be exact. My approximate expression does not become exact after any finite time. For one thing, it is too mathematically smooth; in the exact expressions, given above, only the first n time differentials are continuous at the point $t=(n+1)d/v$.

If γ and μ are zero, but not δ, the exact expressions above remain valid on changing γ to δ/α, and v to v_-.

If neither γ nor δ is zero, but only μ, the exact expression, in the form corresponding to (1), contains two p's, say p_1 and p_2, the roots of

$$1+\frac{\delta}{\psi}+\frac{\alpha\gamma}{\phi}=0,$$

where $\psi=\alpha-(p/v_-)$ and $\phi=\alpha-(p/v)$. When t is less than d/v and is in the range from nd/v_- to $(n+1)d/v_-$ the expression for $I_-(0,t)$ is

$$I_0\sum_{m=0}^{n}\left\{\left(\frac{\delta}{\alpha}\right)^m e^{m\alpha d}\left(1+\gamma+\frac{\delta}{\alpha}\right)^{-(m+1)}+\sum_i\frac{\delta^m(vv_-)^{m+1}}{m!}\left[\frac{d^m}{dp^m}\frac{\psi\phi^{m+1}e^{pt+m\psi d}}{p(p-p_2)^{m+1}}\right]_{p=p_1}\right\} \quad (3)$$

where $\sum_i$ means a summation over an interchange of the suffixes 1 and 2. Thus, when t is less than d/v_-, it is

$$I_0\left\{\left(1+\gamma+\frac{\delta}{\alpha}\right)^{-1}+\frac{vv_-}{p_1-p_2}\left(\frac{\psi_1\phi_1 e^{p_1 t}}{p_1}-\frac{\psi_2\phi_2 e^{p_2 t}}{p_2}\right)\right\},$$

and so forth.

From such expressions as (2) and (3) the values of $I_-(x,t)$ and $I_+(x,t)$ are given by the formulas

$$I_-(x,t)=e^{\alpha x}I_-\left(0,\,t-\frac{x}{v_-}\right);$$

$$I_+(x,t)=\int_x^d \alpha I_-\left(0,\,t+\frac{x}{v_+}-\frac{x'}{v}\right)e^{\alpha x'}dx',$$

in which $I_-(0,T)$ is to be set equal to zero at T negative.

There are several ways of deriving the exact expressions, either in the form such as (2) and (3) or in the form given in my previous paper. If integration in the complex plane is employed the calculation of $I_-(0,t)$ depends on evaluating $\int_C\{e^{pt}f(p)/p\}dp$ where C is an infinite semicircle on the right of the imaginary axis, and $f(p)$ is

$$\left\{\left(1+\frac{\delta}{\psi}+\frac{\alpha\gamma}{\phi}\right)-\left(\frac{\delta}{\psi}e^{\psi d}+\frac{\alpha\gamma}{\phi}e^{\phi d}\right)\right\}^{-1},$$

which has an infinite number of poles. Integrating in the usual way yields the expression for $I_-(0,t)$ given in the section called Part 1 in my previous paper. The alternative form given in the present paper may be obtained by making a binomial expansion of $f(p)$ on the contour C and deleting all terms which, at a given t, will (owing to their character at infinity) give zero integrals. We thus have to integrate a finite number of terms, each having only a few poles. Expressions such as (2) and (3) are obtained.

The expressions such as (2) and (3) are chiefly of interest in the case of a discharge which is highly overvolted, so that even during the first few n's the current amplification becomes very great, and thus of experimental interest. These exact expressions, containing only a few terms, are of value in such cases. But whatever the voltage the form given in my previous paper is more convenient throughout the subsequent growth. That form specifies a definite value of B, say B_0, in the approximate formula, and the correction terms which must be added to make the expression

exact. A few remarks concerning it may be useful to experimentalists. To visualize the nature of B we may write $A-Be^{\lambda t}$ as $(I_0/P)\{1-e^{\lambda(t+\Delta)}\}$. Consider, for example, the case in which γ and μ are zero. Setting Δ zero makes this approximate expression for $I_-(0,t)$ smaller than the exact value at all positive t's (as may be seen from the nature of the error in the initial distribution); on the other hand setting Δ equal to d/v_- makes the expression too large at all positive t's. If the magnitudes of A and αd are large compared with unity, B_0 makes Δ about $d/2v_-$; (the B which Bandel assumes makes Δ nearly d/v_-). In using the exact formula to estimate whether, at a given t, the correction terms are of any importance, there is no need to insert exact values of the complex λ's. It would be sufficient, in this case of γ and μ zero, to write $\lambda_{n\pm}=\bar{\lambda}_n\pm(2\pi i n v_-/d)$ where $\bar{\lambda}_n=\lambda_0-(2\pi^2n^2v_-/\alpha^2d^2)$ at small n's and tends to $-(v_-/d)\log(2\pi n/\alpha d)$ at very large n's. λ_0 is the real λ, which appears in the B_0 term. Thus the correction terms in $I_-(x,t)/I_0$ are approximately

$$\sum_{n=1}^{\infty}\frac{1}{\pi n}\exp\{\bar{\lambda}_n[t-(x/v_-)]\}\cos\left\{\frac{2\pi n v_-}{d}\left(t-\frac{x}{v_-}\right)-\theta_n\right\},$$

θ_n being the angle, less than π, whose tangent is $2\pi n v_-/\bar{\lambda}_n d$.

In using the accompanying formula[1] for $I_+(x,t)$, it is convenient to hasten the convergence at the smaller values of t by subtracting a series which is zero; for example, so long as $v\{t+(x/v_+)\}$ is less than d we may evidently subtract $I_+[v\{t+(x/v_+)\},v\{t+(x/v_-)\}/v_-]$ that is, we may replace the d's in the numerators of the expression for $I_+(x,t)$ by $v\{t+(x/v_+)\}$. If the formula is used without a procedure of this type, the case of γ zero must be regarded as the limiting case of γ very small.

If the distributions of electrons and ions between the plates at time zero are not negligible, the expressions called Part (ii) in my previous paper are required.

Note added in proof.—For the case in which there is only one regenerative process, an expression equivalent to the alternative form of the exact solution has been obtained in a recent paper by P. L. Auer [Phys. Rev. 98, 320 (1955)]. His expression (3.17), which he proposes for practical use, is an approximate formula for $I_-(0,t)$ at integral values of n (with $n=1$ at $t=0$) for the case of γ zero and $\delta d\ll 1$; but he underestimates the error in this expression, at large n's in the case of $V>V_s$. The expression differs from my approximate formula $A-B_0e^{\lambda t}$ in two ways. Firstly, λ has been replaced by a different quantity; it is practically the value, say λ_∞, which λ assumes if αd is increased to infinity while A is kept unchanged; and secondly, B_0 has been changed to a value which makes the expression practically correct at time zero. It may be shown that λ_∞/λ is approximately $\alpha d/(\alpha d+1)$ if αd is fairly large compared with unity (values about ten are common in the experimental work). Thus, if $V>V_s$ his alteration in λ makes his expression (3.17) too small at large n's, the true current exceeding it by a percentage which increases without limit at very large n's. It would have been better to have retained my approximate formula $A-B_0e^{\lambda t}$, in which, of course, the percentage error diminishes to zero at large n's, when the omitted oscillatory terms (the terms with complex λ's) have died out relative to the retained terms.

The physical reason for the accuracy of Auer's formula at small integral n's may be seen by imagining αd increased to a very large value without changing A. Then, owing to the shape of the curve $e^{\alpha x}$, an avalanche generated by an electron at the cathode produces its photoelectrons very suddenly, just before it reaches the anode. Thus the $I_-(0,t)$ curve due to a maintained I_0 becomes practically a staircase during the first few n's, the sudden steep slopes occurring just before the integral n's. If now we diminish αd to a more practical value, we diminish thereby the mean time at which an avalanche, after setting out from the cathode, generates its photoelectrons; we diminish it from d/v_- to about $(\lambda_\infty/\lambda)(d/v_-)$. Thus we may say that we now have a more rapid amplification process, though the effect will not appear appreciably at the first few integral n's. In that range the extra current will appear at intermediate points, as a less steep ascent in the regions approaching the integral n's; but at later times the effect will extend to the integral n's. It should be noted, moreover, that in the range of very small n's a smooth curve drawn through the points given by Auer's formula is only correct at the integral n's, and usually one is not especially interested in those particular points; if a smoothed curve of current growth is required it would have been better to have retained my approximate formula $A-B_0e^{\lambda t}$ even in this range of small n's. For example, in the extreme case of $\alpha d\gg 1$, which makes the true curve almost a staircase in this range, Auer's smooth curve is like an inclined (and rather curved) rod resting on the stairs; its error never changes sign. The curve drawn from my approximate formula intersects each stair, giving an error alternately positive and negative, and is a good smoothed approximation.

F' and K Centers in Additively Colored KCl Crystals

FELIX E. GEIGER, JR.
United States Naval Research Laboratory, Washington, D. C.
(Received April 11, 1955)

Measurements are reported on the $F-F'$ process, and the K band in additively colored KCl. The crystals are colored with distilled potassium, and subjected to repeated cycles of optical bleach with F light (5450 A) at 170°K and thermal bleach of the F' centers at room temperature. It is found that the optical quantum efficiency of conversion of F to F' centers varies in two ways: (a) The quantum efficiency varies significantly for different specimens of colored KCl all cleaved from the same one-inch cube of KCl grown by the Harshaw Company. (b) The optical quantum efficiency drops significantly after each cycle of F-center bleach and subsequent thermal bleach of the F' centers. Ultimately, the quantum efficiency appears to decrease to a point where the crystal is essentially unaffected by F light at 170°K. K centers are influenced to a lesser degree, if at all, than F centers by optical bleaching in the F band at 170°K. All absorption measurements are made in the visible part of the spectrum.

INTRODUCTION

THE formation of F' centers was thoroughly investigated by Pick[1] in additively colored crystals of KCl, KBr, and NaCl. Pick established that the optical quantum efficiency[2] for the conversion of F centers into F' centers in KCl is a maximum at approximately 190°K with a value of 2.0. For NaCl and KBr crystals, the quantum efficiency obtained by him is considerably less.

This report grew out of an investigation of F-band bleaching at low temperatures in additively colored KCl. We noticed in the course of these experiments that the F'-center formation at temperatures of approximate maximum quantum efficiency, i.e., 170°K, was accompanied by several effects not found by Pick. A systematic investigation of F'-center formation and attendant effects in the presence of R and presumably M and N centers was therefore undertaken. In addition, this work yielded new data on the relative behavior of F and K bands at low temperatures.

EXPERIMENTAL PROCEDURE

Figure 1 is a schematic diagram of the apparatus used in the additive coloring process. The glass vial, containing half a dozen pieces of KCl, was outgassed at

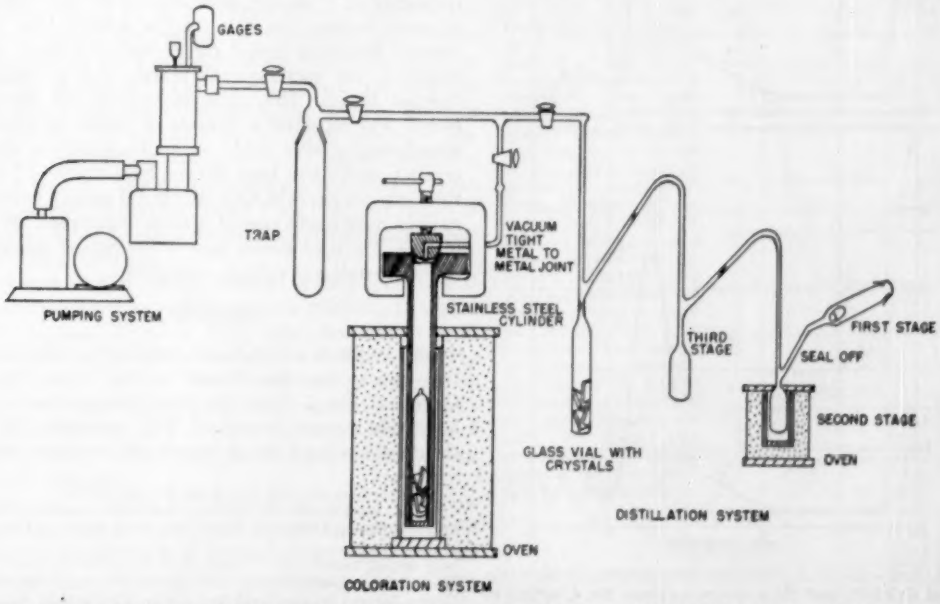


FIG. 1. Schematic diagram of the apparatus used in the distillation of potassium and the additive coloring of KCl crystals.

[1] H. Pick, Ann. Physik **31**, 365 (1938); **37**, 421 (1940).
[2] The optical quantum efficiency is defined as the ratio of the number of destroyed (or bleached) F centers to the number of absorbed light quanta.

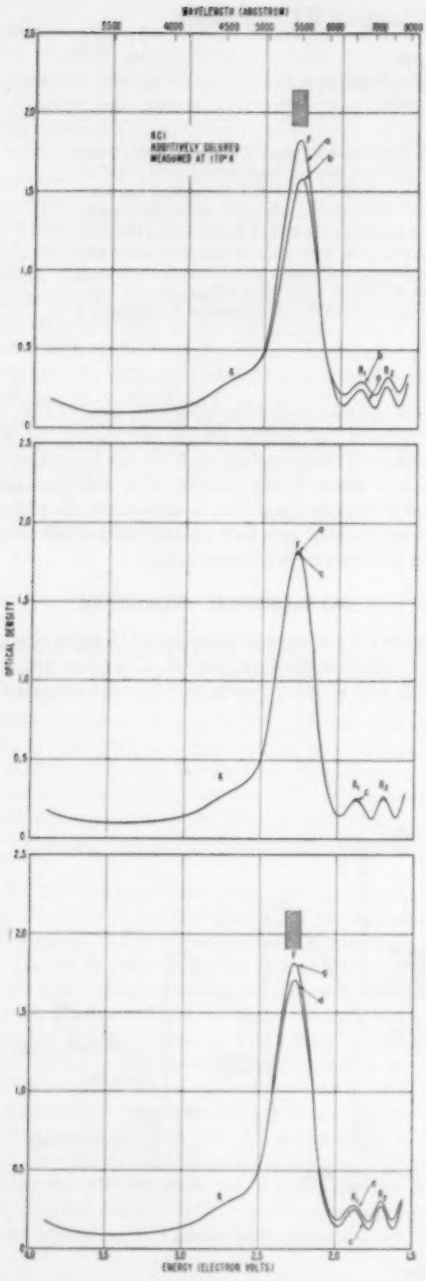


FIG. 2. A sequence of absorption measurements on additively colored KCl 0.017-inch thick, crystal specimen No. 4, subject to repeated cycles of optical bleach at 170°K with *F*-band light and thermal bleach at room temperature in the dark. The absorption measurements were made at 170°K. Curves (a), (b), (c), and (d) correspond to successive stages of optical and thermal bleaching. Curve (a), no optical or thermal bleach; Curve (b), after two hour optical bleach with *F* light (5450 A) at 170°K; Curve (c), after 12-hour thermal bleach of *F'* centers at room temperature. Curves (a) and (c) are shown together to illustrate the degree of recovery of the *F* band after the thermal bleach; Curve (d), after two-hour optical bleach in the *F* band. Curves (c) and (d) are plotted together so that the *F* bands may be easily compared before and after the optical bleach. The shaded area represents the wavelength region covered by the *F* light during bleaching.

415°C for about twelve hours. The potassium was heated in each stage of the distillation process for twenty-four hours at 360°C and 10^{-4} mm Hg, and transferred to the glass vial, which was sealed off and put into the stainless steel container. The steel cylinder was evacuated to prevent collapse of the glass tube at high temperatures. All crystals used in this experiment were heated for twenty-eight hours at 477°C, and quenched by quickly submerging the cylinder in ice water. Absorption measurements were made on crystal sections approximately $\frac{3}{8}$ by $\frac{3}{8}$ by 0.017 inch, which were cleaved from $\frac{1}{2}$ by $\frac{1}{2}$ by $\frac{1}{8}$ inch pieces of colored KCl. The crystals were obtained in one-inch cubes from the Harshaw Company. Optical measurements and *F*-band irradiations were made with a Cary automatic recording spectrophotometer in the wavelength range 3000–8000 A. The apparatus and techniques used in the low-temperature absorption measurements are described in detail elsewhere.[3]

The experimental procedure was as follows: The crystal was illuminated in the *F* band (5450 A) at 170°K for two hours at a slitwidth of 1.5 mm (250 A). The area of irradiation was limited by a $\frac{5}{16}$-inch aperture. Absorption curves were of course taken at 170°K before and immediately after the optical bleach. The *F*-band bleach was followed by a thermal bleach of the *F'* centers at room temperature and in the dark. The length of the thermal bleach was varied, with a minimum of twelve hours. This should have been sufficient to destroy all *F'* centers, since the half-life of *F'* centers at room temperature is but a few hours.[4] After this thermal bleaching period the crystal was again subjected to the same optical bleach, and subsequent thermal bleach. This cycle of optical and thermal bleach was repeated a number of times. Absorption measurements were made on approximately a dozen crystals, which had been cleaved from different $\frac{1}{8}$-inch thick colored pieces of KCl. All optical density determinations were made against a clear comparison crystal of KCl. The light source used in the optical bleaching of the crystals was voltage-controlled.

RESULTS

The results can best be discussed with reference to two typical examples shown in Fig. 2 and Fig. 3. Curve (a), Fig. 2, shows the absorption spectrum of an additively colored crystal of KCl (specimen No. 4) clearly exhibiting *F* and *R* centers and to a lesser extent

[3] F. E. Geiger, Jr., Rev. Sci. Instr. **26**, 383 (1955).

[4] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, New York, 1940), p. 130.

TABLE I. Optical densities of F and R bands in additively colored KCl, crystal specimen No. 4, after repeated optical bleach in the F band, and subsequent thermal bleach of F' centers at room temperature. Irradiation in the F band was carried out at 5450 A, at 170°K with a 1.5 mm (250 A) slit for two hours. The thermal bleach was in the dark at room temperature and of varying length. All absorption measurements were made at 170°K. Thickness of crystal 0.017 inch. See Fig. 2, Curves (a), (b), (c), and (d).

F-band density	% reduction of F centers after optical bleach	R_1-band[a] density	R_2-band[a] density	Remarks
1.82		0.24	0.27	Initial absorption measurement
1.58	13.2	0.29	0.32	Absorption after optical bleach
1.81		0.25	0.27	Absorption after thermal bleach of 12 hours
1.71	5.5	0.27	0.30	After optical bleach
1.77		0.25	0.27	After 12-hour thermal bleach
1.70	3.9	0.27	0.29	After optical bleach
1.78		0.23	0.25	After 84-hour thermal bleach
1.71	3.9	0.25	0.28	After optical bleach
1.77		0.23	0.24	After 36-hour thermal bleach
1.71	3.4	0.25	0.27	After optical bleach
1.76		0.21[b]	0.22[b]	After 300-hour thermal bleach
1.67	5.1	0.23	0.24	After optical bleach

[a] The R bands are superimposed on the F' band, and the apparent growth of the R bands is of course due to F'-center formation.
[b] The decrease in R centers is almost certainly attributable to thermal bleaching.

K centers. Curve (b) is the absorption curve immediately after the two-hour F-light (5450 A) bleach at 170°K. It shows a 13.2 percent reduction in the number of F centers, and a fairly uniform increase of absorption on the low-frequency side of the F band due to the very broad F' band. Curve (c), obtained after a twelve-hour thermal bleach of the F' centers at room temperature, is compared with Curve (a) on the same graph. It shows complete recovery of the F centers within the accuracy of the instrument. The crystal is now subjected to another F-light bleach at 170°K, the result is shown in Curve (d). We now find that an equal number of absorbed light quanta effects only a 5.5 percent reduction in F centers. This means the quantum efficiency has dropped by 58 percent. Table I is a complete record of F- and R-band optical densities for six optical and five thermal bleaches.

Figure 3 illustrates the behavior of the colored crystal specimen No. 5, subjected as before to alternate optical and thermal bleaches. This time, however, the F' centers were bleached thermally for 60 hours after the first F-band bleach. The F band has recovered now only 95 percent of its original number of centers [Curves (e) and (g)], and we note again a reduction of 44.6 percent in the quantum efficiency for the second optical bleach.

Figure 2, Curve (b) and Fig. 3, Curve (f), show another curious behavior of the optical quantum efficiency. The number of F centers destroyed after the first F-light bleach varies considerably for different crystal specimens although the number of quanta absorbed by each crystal is essentially the same. Results on a few crystal specimens are summarized in Table II. K centers appear on all absorption curves, though un-

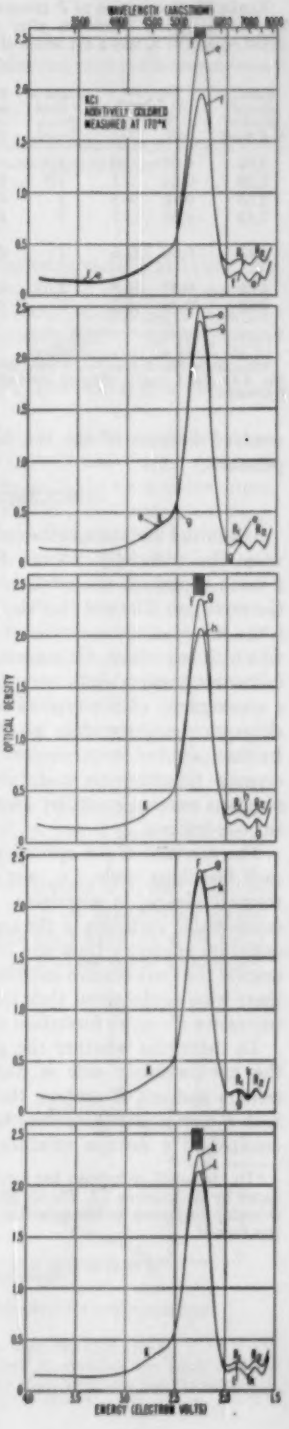


FIG. 3. A sequence of absorption measurements on additively colored KCl 0.017 inch thick, crystal specimen No. 5, subject to repeated cycles of optical bleach at 170°K with F-band light and thermal bleach at room temperature in the dark. The absorption measurements were made at 170°K. Curves (e), (f), (g), (h), (k), and (l) represent successive stages of optical and thermal bleaching. Curve (e), no optical or thermal bleach; Curve (f), after two-hour optical bleach with F-band light (5450 A) at 170°K; Curve (g), after 60-hour thermal bleach of F' centers at room temperature. Curves (e) and (g) are shown together to show the degree of recovery of the F band after the thermal bleach. Curve (h), after two-hour optical bleach with F light (5450 A) at 170°K. Curves (g) and (h) are plotted together to show the decrease in the number of F centers due to the F light bleach. Curve (k), after 36-hour thermal bleach at room temperature. Curves (g) and (k) are shown together for comparison; Curve (l), after two-hour optical bleach in the F band at 170°K. Curves (k) and (l) should be compared with Curves (e) and (f). They show clearly the decrease in the number of bleached F centers after three optical bleaches and two thermal bleaches, even though the number of quanta absorbed in each optical bleach is the same. The shaded area represents the wavelength region covered by the F light during bleaching.

TABLE II. Destruction of F centers in additively colored KCl in various crystal specimens after irradiation in the F band (5450 A) at 170°K and a slit width of 1.5 mm (250 A).

Initial optical density of F band	Decrease in F-band density ΔD	%	Length of F-band bleach (hours)	Thickness of crystal (inch)	Colored crystal specimen No.
2.04	0.04	2.0	$2\frac{1}{4}$[a]	0.014	1[b]
1.28	0.04	3.1	$2\frac{1}{4}$[a]	0.014	2[b]
2.53	0.16	6.3	2	0.018	3
1.82	0.25	13.7	2	0.017	4 See Fig. 2, Curves (a), (b)
2.48	0.56	22.6	2	0.017	5 See Fig. 3, Curves (e), (f)
1.86	0.27	14.5	2	0.016	6
3.01	0.34	11.3	2	0.019	7

[a] Light source not voltage-stabilized.
[b] Specimens No. 1 and No. 2 were obtained from the same, crystals No. 3 to No. 7 from a different one-inch cube grown by the Harshaw Company.

resolved because of the relatively high crystal temperature.

CONCLUSIONS

The initial variation of the optical quantum efficiency when the additively colored KCl is bleached in the F band is perhaps not quite so unexpected considering the results on KBr and NaCl by Pick.[1] More astonishing is the extent of the variation of the quantum efficiency, which is an order of magnitude.[5] These variations follow no predictable pattern and do not appear to be a consequence of the type of coloring procedure used, although definitive data on this point are lacking. In fact, control experiments were carried out with crystals colored with undistilled potassium, but the program was not extensive enough to reach any clear-cut conclusions.

The decrease of the optical quantum efficiency after each bleaching cycle, i.e., one optical bleach and one thermal bleach, is surprising. Here again we note a considerable variation in the amount of decrease in the quantum efficiency from one crystal to the other. But despite the quantitative individuality of the specimens, there is no doubt about their qualitative behavior after successive F'-center formation and thermal destruction.

To determine whether the growth of the bands on the low-frequency side is due predominantly to F' centers and not R centers, the crystal was irradiated with F light at 230°K for two hours. The optical bleach resulted in a definite enhancement of the R centers which remained undiminished after a twelve-hour thermal bleach at room temperature. Therefore, no R centers are formed at 170°K, but at 230°K the R center formation predominates and few F' centers are stable enough to be measured.

It is not clear whether the partial recovery of the F band after optical bleaching is connected with the successive decrease of the quantum efficiency. Apparently, it is not [see Fig. 2, Curves (a) and (c)]. However, the electrons freed by the partial F-band recovery ought to coagulate with vacancy clusters evidenced in the absorption spectrum by increased absorption of existing bands or new bands. Neither could be observed in the wavelength range considered.

The behavior of the ratio of F to K centers is of considerable theoretical interest.[4,6,7] Figure 3, Curves (e) and (f) show a reduction in the number of F centers of 22.6 percent after the usual optical bleach. A proportionate reduction of K centers expected on the basis of Mott and Gurney's theory[4] would reduce the K-band density to 0.30. If we assume that the growth of the F' band contributes roughly 0.06 density units to the K band, then the K band after bleaching should have a density of approximately 0.36. Figure 3, Curve (f) shows that the actual reduction of K centers was much less than 22.6 percent.

Pick[1] obtained the shape of the F' band by subtracting the absorption curves before and after the F-light bleach. Subtracting curve (e) from (f) of Fig. 3 one obtains instead of a broad band with a maximum at 1.80 ev (F' band), three broad bands with peaks at 1.97 ev, 1.79 ev, and 1.64 ev. It appears as if the F' band has superposed on it three weak bands. The control experiments at 230°K, which gave an increase in R centers after F-light exposure did not show these three bands.

ACKNOWLEDGMENTS

I am indebted to Dr. H. Friedman for having suggested a reinvestigation of additively colored alkali halides, and to Dr. D. Dexter, Dr. H. Etzel, and Dr. C. Klick for various helpful discussions on F centers. I am grateful to Dr. R. Tousey and Mr. L. Dunkelman for generously allowing me to use their Cary spectrophotometer almost uninterruptedly for more than a year. I am happy to acknowledge Dr. M. A. Garsten's help in discussing my manuscript, and Dr. Jordan Markham's discussion on new F'-center results.

[5] Dr. James H. Schulman has kindly called my attention to a paper by A. Smakula [Z. Physik **59**, 603 (1930)] in which considerable variations in the quantum efficiency of x-rayed NaCl was found.

[6] H. W. Etzel and F. E. Geiger, Jr., Phys. Rev. **96**, 225 (1954).
[7] F. E. Geiger, Jr., Phys. Rev. **97**, 560 (1955).

Hall Effect in Permalloys*

SIMON FONER†
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received March 18, 1955)

Measurements of the Hall effect for two permalloys of approximately 25 percent and 55 percent Fe in Ni are reported and compared with recent results for some Co-Ni and Fe-Ni alloys. Qualitative aspects concerning the band structure of these Fe-Ni alloys are briefly discussed.

THE Hall effect in ferromagnetic materials may be written as

$$e_H = R_0 B + (R_1 - R_0) 4\pi M, \quad (1)$$

where e_H is the Hall potential difference per unit current per unit thickness, R_0 the ordinary Hall effect, B the magnetic induction, R_1 the extraordinary Hall effect, and M the magnetization. At sufficiently high magnetizing fields, the term in M is independent of B in Eq. (1), and R_0 may be determined unambiguously. Recent results of Schindler and Pugh,[1] for the Cu-Ni alloys, and those of Foner and Pugh[2] (referred to as A) for the Co-Ni alloys, have demonstrated that the effective number of conduction electrons per atom n^* in Eq. (2), are in close agreement with values expected from simple band models. The relation between R_0 and n^* is given by

$$R_0 = 1/n^* Nec, \quad (2)$$

where e is the absolute value electronic charge, c the velocity of light, and N the number of atoms per cc. Recently Karplus and Luttinger[3] have given a theory for the term in Eq. (1) which is proportional to M. We are concerned here mainly with the value of R_0 which is related to the band structure of the metals.

The data for the two permalloys are shown in Fig. 1 and the high-field portions are shown in Fig. 2. The Ettingshausen correction above saturation is estimated to be less than 1 percent, if one uses the measured Ettingshausen effect and thermocouple data from other sources. The permalloy No. 1 sample[4] was machined from a forged bar before annealing in vacuum at 620°C for two hours and slowly cooling to room temperature while the permalloy No. 2 sample[5] was machined from a small slab and annealed in vacuum at 1150°C for four hours, then slowly cooled to room temperature. Both samples were approximately 6 cm long, 1 cm wide, and 0.100 cm thick. Measurements of the Hall effect were made by a dc method up to 30 kilogauss using the experimental arrangement and techniques described in A.

The values of R_0 *vs* electron concentration for these two permalloys and data of Smit and Volger[6] (referred to as S and V) are compared with the room temperature results for Co-Ni of A in Fig. 3. The deviations of the S and V data from the previously presented Co-Ni data are somewhat difficult to explain unless insufficient magnetic fields were used by those investigators. That this possibility is likely is based on the following reasoning. From Eq. (1) we see that if the coefficient of the M term is large, then small variations of M would be reflected as large variations of e_H.[7] Only when the coefficient of M is zero (*a* special case) or when M is strictly independent of B, can R_0 be determined accurately. This latter condition is closely satisfied well above the usual "technical saturation" as shown in Fig. 2, where it can be seen that fields over 16 kilogauss were needed to establish the value of R_0. S and V used fields up to 14 kilogauss.[6]

Examination of the low-field data in A (see Fig. 2 of A) shows that the S and V Co-Ni data deviate in the manner expected if the fields used were not sufficiently high to eliminate the term in M. One might also expect

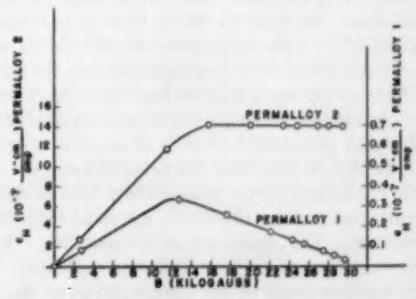


FIG. 1. Comparison of Hall effect for two permalloys.

* The research in this document was supported by the Office of Naval Research.

† Presently at Lincoln Laboratory, Massachusetts Institute of Technology.

[1] A. I. Schindler and E. M. Pugh, Phys. Rev. **89**, 295 (1953).

[2] S. Foner and E. M. Pugh, Phys. Rev. **91**, 20 (1953). This paper will be referred to as A.

[3] R. Karplus and J. M. Luttinger, Phys. Rev. **95**, 1154 (1954).

[4] This material, kindly furnished by Westinghouse Electric and Manufacturing Company, was used for neutron diffraction data reported by J. E. Goldman, Revs. Modern Phys. **25**, 108 (1953). Analysis by Pgh. Testing Laboratory—analysis of impurity not available. Nominal composition by weight of this sample: 71.8 percent Ni, 25.9 percent Fe, 2.4 percent impurities.

[5] Nominal composition of this sample kindly furnished by Dr. V. E. Legg, of the Bell Telephone Laboratories: 45 percent Ni, 54.5 percent Fe, 0.5 percent Mn and small percentages of impurities such as cobalt and silicon.

[6] J. Smit and J. Volger, Phys. Rev. **92**, 1576 (1953).

[7] In fact, this term may be used to examine very small variations in M by means of the Hall effect assuming Eq. (1) holds—see S. Foner, Phys. Rev. **95**, 652(A) (1954).

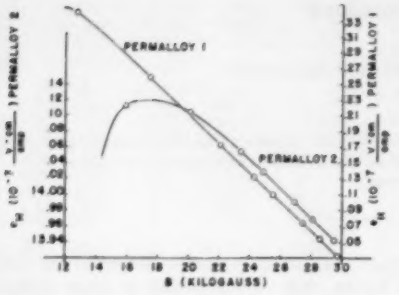


FIG. 2. Comparison of ordinary Hall effect for two permalloys at high fields.

the iron nickel data of S and V to behave in the same manner for insufficient fields if they had effects similar to the Co-Ni alloys. If the above objection is valid, the apparent R_0 observed by S and V would vary strongly with temperature in a somewhat unpredictable manner since both the coefficient of M, and M as a function of H, may depend strongly on temperature. Possibly this explains the larger variation of the S and V Cu-Ni data as a function of temperature compared to that of Cohen and Pugh.[8] The data for R_0 given in A as well as the permalloy data presented here were taken at sufficiently high fields (up to 30 kilogauss) to obtain a strictly linear variation of e_H with B at the high fields.

Since the Fe-Ni and Co-Ni alloys are face-centered cubic structures for high Ni concentrations and furthermore their saturation moment *versus* electron concentration are identical over the same range, one might expect a similar band structure. The close agreement of R_0 for the 25 percent Ni ($R_0=-18.8\times10^{-12}$ v-cm/amp-gauss) and 50 percent Co−50 percent Ni ($R_0=-19.6\times10^{-12}$ v-cm/amp-gauss) seems to justify this point of view. The difference between the 55 percent Fe permalloy ($R_0=-18.7\times10^{-12}$ v-cm/amp-gauss) and pure Co is not surprising, since they do not have quite the same electronic concentration and, furthermore, their phases are different. A two-band conduction model for 4*s* and 3*d* electrons presented in A[2] to explain the detailed variation of R_0 with composition for the Co-Ni alloys encountered difficulties when it showed that the number of *s* electrons per atom must be less than 0.35. Recently Pugh[9] has devised a four-band scheme composed of both parallel and antiparallel 4*s* and 3*d* bands to overcome this difficulty. It describes the detailed behavior of R_0 for the Cu-Ni and Ni-Co alloys as functions of electron concentration and for the Cu-Ni alloys as functions of temperature. The construct of antiparallel *s* bands allows these bands to have different mobilities. This model ascribes most of the conductivity to the high mobility of those *s* electrons with parallel spins, which cannot be scattered in the *d*-band because there are no vacancies in the parallel *d*-band. The description for the Co-Ni alloys now proceeds as in A. We assume only vacancies in antiparallel 3*d* band in this region. For high Ni concentrations we have both 3*d* hole conduction and 4*s* electronic conduction. As the Co concentration increases, we expect the 3*d* density of states to increase, resulting in a lower 3*d* mobility and a decreasing R_0. The maximum density of states is thus expected near 40 percent Co in Ni. Near 70 percent Co in Ni a decrease in density of states of the 3*d* band is expected with a resulting electronic conduction from this antiparallel 3*d* band. Throughout this region the 3*d* band hole or electron conduction competes with the 4*s* electronic conduction. The saturation moment data implies that about 0.6 to 0.7 electron/atom are available over the Ni-Co alloy range. If in the region where R_0 is constant, the mobility of the antiparallel 3*d* electrons is low because of the high density of states and furthermore the mobility of the antiparallel 4*s* electrons is also low because of scattering in the 3*d* band, then, only the parallel 4*s* electrons contribute appreciably to the Hall effect. Thus, one expects to observe 0.35 electron/atom from the Hall effect which corresponds closely to the observed $R_0\sim19\times10^{-12}$ v-cm/amp-gauss. The similarities between the data on the Fe-Ni alloys with low Fe concentrations[10] and that on Co-Ni alloys of low Co content suggest that the same model may be used with minor modifications for both of these systems at the same electron concentrations.

For the 25 percent Fe−75 percent Ni permalloy, $R_1=33.9\times10^{-12}$ v-cm/amp-gauss at 16°C whereas for the 55 percent Fe−45 percent Ni permalloy $R_1=875\times10^{-12}$ v-cm/amp-gauss at 28°C. This large variation in R_1 is also apparent on comparing Figs. 1 and 2 with Eq. (1). Although in this case R_1 increases with ρ as predicted by the Karplus and Luttinger[3] theory, this theory is probably not applicable to variations in R_1 and ρ on alloying. For example, when ρ is increased by

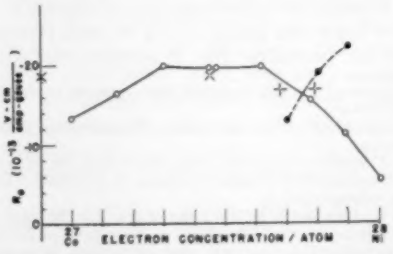


FIG. 3. The ordinary Hall effect *versus* electron concentration obtained by recent investigations of Co-Ni and Fe-Ni alloys. ●—Co-Ni data of Smit and Volger. ○—Co-Ni data of Foner and Pugh. +—Fe-Ni data of Smit and Volger. ×—Fe-Ni data presented in this paper.

[8] The author is indebted to Dr. Philip Cohen and Professor E. M. Pugh for making their data available previous to publication.

[9] E. M. Pugh, Phys. Rev. **97**, 647 (1955).

[10] R. M. Bozorth, *Ferromagnetism* (D. Van Nostrand Company, Inc., New York, 1951), Figs. 5 and 6.

adding Co to pure Ni, R_1 decreases rather than increases.

It appears that investigations of the Fe-Co alloys and the Fe rich Fe-Ni alloys would be interesting in examining the detailed variations of R_0 as a function of electron concentration.

I would like to thank Professor Emerson M. Pugh and Professor J. E. Goldman for their comments and interest in this work. I am also grateful to the Bell Telephone Laboratories and Westinghouse Research Laboratory for the materials used in this investigation and the Office of Naval Research for support of this work.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Long-Wavelength Neutron Transmission as an Absolute Method for Determining the Concentration of Lattice Defects in Crystals*

J. J. ANTAL† AND R. J. WEISS, *Ordnance Materials Research Office, Watertown Arsenal, Watertown, Massachusetts*

AND

G. J. DIENES, *Brookhaven National Laboratory, Upton, Long Island, New York*

(Received April 4, 1955)

The scattering of long-wavelength neutrons by lattice defects has been used to measure in an absolute way the number of isolated interstitial atoms and vacancies introduced in graphite by reactor irradiation. For an exposure of 1.1×10^{20} neutrons/cm² the fraction of displaced atoms was found to be 0.0263. This value is more accurate than present theoretical estimates largely because of uncertainties in exposures. Within this limitation the experimental value is in excellent agreement with that derived from Seitz's theory.

I. INTRODUCTION

IT is well known by now that an appreciable number of lattice defects, presumably primarily interstitial atoms and vacant lattice sites, can be introduced into a solid by high energy particle bombardment.[1] The number of defects introduced by such bombardment has been estimated theoretically by Seitz.[2] Experimentally only indirect methods have been available for determining the concentration of lattice defects. The basic difficulty is that the dependence of changes in various physical properties due to the introduction of defects is not well enough understood theoretically to furnish reliable methods for an absolute (in contrast to relative) determination of defect concentration. The purpose of this paper is to describe a method, based on the scattering of long-wavelength neutrons by the lattice defects, which has been used successfully to determine the concentration of imperfections introduced by exposure in the Brookhaven reactor.

Optical methods (for color centers) and residual resistivity measurements (for defects not present as color centers) have been the most successful methods so far employed for determining the concentration of various defects in crystals. Both of these, of course, depend either on absorption or scattering. In *F*-center work, it is possible to obtain absolute numbers for the color centers by optical absorption methods. Residual restistivity measurements have not yielded absolute numbers because the scattering cross sections for conduction electrons are not known with sufficient theoretical accuracy.

It will be shown that long-wavelength neutrons can be used very similarly to conduction electrons and their scattering by vacancies and interstitials may be measured. Neutrons of sufficiently long wavelength are scattered isotropically by isolated point defects and the scattering can be measured when crystalline effects (Bragg scattering) are absent. Babinet's principle may be applied under such conditions and, therefore, vacancies and interstitials scatter in exactly the same manner. The cross section for this nuclear type of scattering is accurately known from other measurements in contrast to scattering of conduction electrons by the atomic electrons. Thus, if the scattering from these defects is measurable an absolute method is at hand for determining their concentration. In the theoretical section the above scattering process is described mathematically. The experimental results and their discussion are given in Secs. III and IV.

II. THEORY

Let long-wavelength neutrons be incident on a crystal in the wavelength region of several angstroms, i.e., the energy region of 0.001 ev. For such extremely low-energy neutrons the solution of Schrödinger's equation is the sum of an incident plane wave and a radially scattered wave. The scattered radiation at a

* Under contract with the U. S. Atomic Energy Commission.

† Guest scientist at Brookhaven National Laboratory.

[1] For recent reviews, see: J. C. Slater, J. Appl. Phys. **22**, 237 (1951); G. J. Dienes, Ann. Rev. Nuc. Sci. **2**, 187 (1953).

[2] F. Seitz, Discussions Faraday Soc. No. 5, 271 (1949).

large distance is

$$\psi=\sum_p ae^{i\mathbf{K}\cdot\mathbf{r}_p}, \tag{1}$$

where $\mathbf{K}=\mathbf{K}'-\mathbf{K}_0$=phase difference and $\mathbf{r}_p$=vector distance between origin and pth scatterer in the crystal. Let a monatomic crystal contain m point scatterers per unit volume in the form of interstitial atoms and vacant lattice sites and N atoms per unit volume. The scattering is then described by

$$\psi=\sum_{j=1}^{N} ae^{i\mathbf{K}\cdot\mathbf{r}_j}+\sum_{i=1}^{m} be^{i\mathbf{K}\cdot\mathbf{r}_i} \tag{2}$$

where b is negative for the vacancies. The differential cross section for scattering is given by

$$\sigma(\theta)=\psi\psi^*=a^2\sum_{j,j'}^{N} e^{i\mathbf{K}\cdot(\mathbf{r}_j-\mathbf{r}_{j'})}+b^2\sum_{i,i'}^{m} e^{i\mathbf{K}\cdot(\mathbf{r}_i-\mathbf{r}_{i'})} +\text{cross terms}. \tag{3}$$

The primes here distinguish between scatterers of the same kind. The cross terms are strongly peaked in the forward direction and, therefore, do not contribute to the scattering. The first term of Eq. (3) represents Bragg diffraction which under these conditions (long wavelength) is absent. Thus

$$\sigma(\theta)=a^2\sum_{i,i'}^{m} e^{i\mathbf{K}\cdot(\mathbf{r}_i-\mathbf{r}_{i'})}. \tag{4}$$

Equation (4) for $i\neq i'$ is zero because a random distribution of scatterers is assumed. Therefore,

$$\sigma(\theta)=ma^2. \tag{5}$$

The integrated scattering cross section for these scatterers is,

$$\sigma=4\pi ma^2. \tag{6}$$

The cross section per atom is, therefore,

$$\sigma_d=\sigma/N=4\pi a^2m/N=4\pi a^2f, \tag{7}$$

where f is the atomic fraction of scatterers in the material and σ_d is the cross section for scattering by defects alone. The scattering process is fully described, therefore, by $4\pi a^2$ which is the scattering cross section for vacancies and interstitials, and is the bound-atom cross section, σ_b, for the atoms of the crystal.

It is not practical to attempt to measure directly the isotropically scattered neutron intensity. Instead, the attenuation of a long-wavelength neutron beam during its passage through the material is measured in a transmission experiment. There are other sources of attenuation which have to be taken into account and whose cross sections should be very small compared to the cross section for defect scattering. The total attenuation of the neutron beam is determined by a cross section:

$$\sigma=\sigma_a+\sigma_i+\sigma_{dis}+\sigma_d, \tag{8}$$

where σ_a=cross section for absorption (capture), σ_i=cross section for inelastic scattering, σ_{dis}=cross section for disorder scattering other than crystal defects (isotopic, spin, etc.), and σ_d=cross section for defect scattering [Eq. (7)].

In the absence of defects and past the last Bragg cutoff ($\lambda>2d_{max}$), the transmitted intensity, I_s, is given by

$$I_s=I_0\exp[-NX(\sigma_a+\sigma_i+\sigma_{dis})], \tag{9}$$

where I_0=incident intensity, N=number of nuclei per cm^3, and X=path length traversed through the sample. If m defects are present, the transmitted intensity, I_d, is

$$I_d=I_0\exp[-NX(\sigma_a+\sigma_i+\sigma_{dis})-mX\sigma_b], \tag{10}$$

or

$$I_d=I_0\exp[-NX(\sigma_a+\sigma_i+\sigma_{dis}+\sigma_b f)]. \tag{11}$$

A direct comparison of a crystal containing a fraction, f, of defects to a control crystal gives

$$I_d/I_s=\exp(-NX\sigma_b f). \tag{12}$$

The quantity f is then finally calculated from

$$f=\frac{-\ln(I_d/I_s)}{NX\sigma_b}. \tag{13}$$

III. EXPERIMENTS

Graphite was chosen for study because it fulfilled very well the theoretical and experimental requirements. Samples available to us were expected to contain of the order of a few percent defects on the basis of Seitz's[2] theory. This concentration should be easily measurable. It has already been mentioned that the cross section for defect scattering should be much larger than any other in the energy region where the scattering occurs. For graphite,[3]

$$\sigma_b=4.7 \text{ barns},$$
$$\sigma_a+\sigma_i\simeq 0.9 \text{ barn at 8 A},$$
$$\sigma_{spin}=0, \text{ zero spin},$$
$$\sigma_{isotope}=0.$$

σ_{isotop} is negligibly small because of the combination of a low abundance of C^{13} (1.1 percent) relative to C^{12} and a similarity in cross sections, 4.5 b and 5.5 b, respectively. The graphite specimen served also as a neutron filter, as described below, which resulted in a most economical use of the very low intensity available in a long-wavelength beam.

The slow-neutron beam was obtained from the Brookhaven reactor by filtering the thermal-neutron spectrum.[4] This spectrum has a Maxwellian energy distribution peaked near 1 A with a "tail" on the long-

[3] D. J. Hughes, *Pile Neutron Research* (Addison-Wesley Press, Cambridge, 1953), pp. 349–366, 250.
[4] Reference 3, pp. 239–242.

wavelength side. If a polycrystalline material of sufficient length is placed in such a beam, Bragg scattering removes all neutrons from the incident beam except those having

$$\lambda > 2d_{\max}, \tag{14}$$

where $d_{\max}$ is the largest interplanar spacing for which diffraction is possible. Equation (14) is merely the Bragg equation for diffraction for $\theta > 90°$. All neutrons of wavelengths given by Eq. (14) pass through the polycrystalline filter almost unattenuated and constitute a spectrum of long wavelength neutrons for experimental use. In these experiments the graphite specimens were made long enough (approx. 9 in.) to constitute efficient filters by themselves. A typical spectrum of the neutrons transmitted by a 9-in. graphite specimen is shown in Fig. 3 as the "unirradiated" curve. The "cut-off" wavelength is clearly marked by an abrupt increase in transmitted intensity at $\lambda = 2d_{(002)} = 6.70$ A.

Originally it was intended to measure directly the total intensities I_d and I_s transmitted by an irradiated and an unirradiated specimen by placing the graphite in the beam with a $B^{10}F_3$ counter immediately behind it. However, pile irradiation increases the c-axis lattice parameter which determines $d_{(002)}$ in graphite. The larger "cutoff" wavelength decreases the transmitted intensity independent of defect scattering and this effect is very hard to correct for. It was decided, therefore, to obtain a plot of transmitted intensity vs wavelength using a crystal spectrometer. The irrelevant change in intensity could then be disregarded and the transmitted intensity obtained by computing the areas under the curves for $\lambda > 7.30$ A (see Fig. 3). Onset of second-order reflections at about 13.4 A would place a limit on the usable wavelength region but is of no consequence in these experiments since the beam intensity is too weak to be detected past 12 A.

The presence of a large number of pores in graphite broadens the transmitted beam considerably by multiple small angle scattering. Care was taken in the placement of the counter relative to the specimen and in defining the incident beam in order that all small angle scattered neutrons be accepted. Figure 1 shows that this was accomplished with the counter kept adjacent to the specimen during all runs. The spectrometer setup itself is described in Fig. 2.

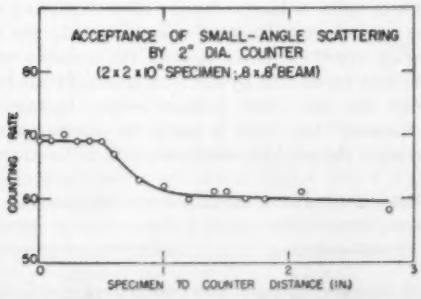


FIG. 1. Recorded transmitted intensity as counter is moved away from the specimen. For the geometry used, loss of small-angle-scattered neutrons does not occur until the separation is ¼ inch.

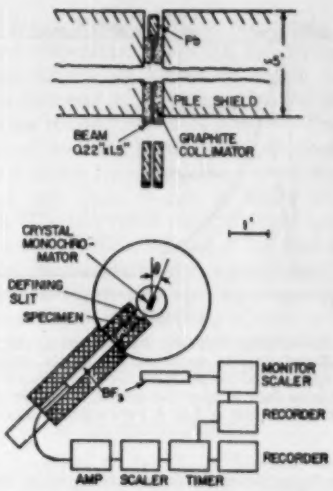


FIG. 2. Crystal spectrometer as employed in these studies. The crystal monochromator is a stack of mica sheets reflecting from the (001) cleavage planes, so that $\lambda = 19.908 \sin\theta$. The defining slit dimensions are 0.81 in.×0.81 in. Six inches of Pb at the collimator provide fast neutron and γ-ray attenuation. A portion of the slow-neutron beam transmitted by the Pb is intercepted by a BF_3 counter to provide monitoring of any spurious fluctuations in incident intensity.

Because the irradiated graphite specimen was not run previous to exposure, an unirradiated specimen of the same type of graphite (high-purity AGHT−*CS*) was selected as a standard. Several samples of this graphite each 2 in.×2 in.×10 in., were taken from various large extruded blocks. The neutron spectrum transmitted by each for various orientations was examined and it was quite evident that neither nonuniformity of the graphite nor preferred orientation were of any consequence in this experiment and no special care in selection and positioning of these large specimens was required. In order that the unirradiated specimen have an absorption in the neutron beam identical to that of the irradiated specimen, the number of atoms per unit volume, N, was made closely the same in both. Since

$$N = (A/M)\rho l,$$

where A is Avogadro's number and M is the molecular weight of graphite, the product of density ρ and length l should be the same in both specimens. The density of the irradiated specimen was obtained from its weight and gross dimensions to give $\rho l = 40.277$ g/cm². It was possible to adjust the length of the standard to $\rho l = 40.188$ g/cm², a ¼ percent difference which does not

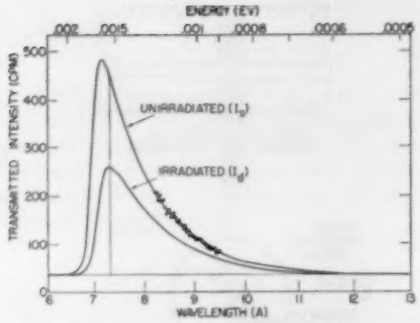


FIG. 3. Slow-neutron intensity transmitted by an irradiated and unirradiated graphite specimen. For clarity, only a typical group of experimental points have been reproduced along one curve to indicate their number and spread. Intensities only to the right of the vertical line at 7.30 A were considered in computing I_u and I_d.

affect the transmitted intensity well within the overall experimental error.

Several spectra were obtained from the spectrometer for the irradiated and standard (unirradiated) specimen run alternately and averaged. Figure 3 shows the resultant transmitted spectra. The areas under each curve give

$$I_d/I_u=0.608.$$

The estimated accuracy of this figure is about ±9 percent. By Eq. (13),

$$f=0.0526,$$

where f is the total fraction of single scatterers, i.e., interstitials and vacancies. The fraction of displaced atoms is, of course, just one-half of this number, namely $f/2=0.0263$.

Annealing experiments were also carried out on another sample of irradiated graphite whose initial defect concentration was smaller. Values of f were determined after various anneals as shown in Table I.

An unirradiated sample was also heated at 900°C for 1 hr. No difference in the transmitted spectra was found before and after annealing. Further, after the 2000°C anneal the sample and the control agreed very closely. The above results show clearly, therefore, that a true irradiation effect was measured in these experiments. The annealing data showed that at or past a 900°C heat treatment the defects could no longer be measured by long-wavelength neutron transmission within the accuracy of the present experiments. Since it is known that many physical properties do not recover completely unless the graphite is heated to about 2000°C, these results can only mean that the major fraction of the remaining defects are present in such a form that they no longer are effective as single scatterers for long-wavelength neutrons. One might expect that pairs of defects or larger clusters should be observable after a partial anneal. The present experiments were not sufficiently accurate over a wide enough wavelength region past 9 A to warrant an analysis for such clusters.

IV. DISCUSSION

An estimate of the number of displaced atoms expected on the basis of Seitz's[2] theory may be obtained from the equations:

$$\tfrac{1}{2}f=(nvt)\times\sigma_s(\Delta\bar{E}/E_0)^1, \quad \Delta\bar{E}=2mME/(M+m)^2, \quad (15)$$

where m = mass of neutron, M = mass of carbon atom, $\Delta\bar{E}$ = energy loss per collision, E_0 = energy required to displace a carbon which will be taken as 25 ev, E = average energy of fast neutron causing displacement, which will be taken as 1 Mev, σ_s = collision cross section for carbon atom = 2.5×10^{-24} cm², and (nvt) = effective total integrated flux of fast neutrons causing displacement (nv is the flux of neutrons per cm² per second).

The number to be used for (nvt) is the most uncertain quantity in these equations since the fast flux and its energy distribution are not known with any accuracy. Our best estimate of the effective (nvt) for the graphite sample used in these experiments is 1.1×10^{20} neutrons/cm².

$\Delta\bar{E}$ is 0.142 Mev and with 1.1×10^{20} neutrons/cm² for nvt, Eqs. (15) give, for the number of displaced atoms,

$$\tfrac{1}{2}f=0.021.$$

The uncertainty in this number resulting from inaccuracy in (nvt) is of the order of 50 percent. There may be a further error because of possible annealing of the specimen although its temperature during irradiation probably has not been above 50°C.

The experimentally determined value of f is known, therefore, with far greater accuracy than any theoretically derived value largely because of uncertainties in the value of (nvt). Consequently, the theory itself cannot be judged critically. The fraction displaced atoms determined experimentally by this method is in excellent agreement with the theory within the limitations mentioned above.

The value of f determined in these experiments may be in error. One reason is that if the defects are aggregated into pairs or larger clusters their scattering will not be equivalent to those of isolated interstitials and vacancies. Another reason is that the inelastic cross section may be altered by the irradiation. It has been assumed that this effect is unimportant. It has also been assumed that there is negligible inhomogeneous distortion in the neighborhood of the defect. An outward

TABLE I. Effect of annealing on defect concentration.

Annealing treatment	f = fraction of single scatterers
None	0.033
250°C for 3 hr.	0.021
500°C for 1 hr	0.011
900°C for 1 hr	0.00
2000°C for 1 hr	0.00

inhomogeneous distortion around the interstitial can be shown to reduce the effective cross section of the interstitial (a similar distortion would increase the cross section of the vacancy). Theoretical estimates of this correction are at present unreliable but are probably not greater than 20 percent.

In principle, the occurrence of pairs is detectable by examining the wavelength dependence of the attenuation. These preliminary experiments are not sufficiently accurate past 9 A to establish the existence of a wavelength dependence. The fact that no wavelength dependence which would be outside experimental error is observable indicates that only a small fraction of the displaced atoms may be present in the form of pairs. More refined experiments will be necessary to establish this point definitely.

The technique described here is applicable, of course, to other materials. The most important limitation is that the absorption cross section has to be small. It appears that with some refinements the number of displaced atoms may be determined in, say, aluminum ($\sigma_a = 0.215$ barn). Experimentation with such a metal calls for low-temperature irradiation and measurement because the defects are known to anneal out well below room temperature. Consequently, the necessary techniques are considerably more involved. Sensitivity may be gained, however, by measuring the total transmitted beam rather than using the spectrometer technique since very little change in lattice parameter is expected. The last cutoff for aluminum appears at $\lambda = 2d_{(111)} = 4.66$ A which, because of the nature of the pile neutron spectrum, also offers a substantial increase in intensity. However, this may be compensated for almost completely by the necessity of using a beam of smaller cross section and an external filter because present low-temperature irradiation facilities limit specimen sizes.

ACKNOWLEDGMENTS

We are indebted to Dr. J. S. Koehler for many stimulating and helpful discussions during the early days of these experiments.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Calculation of the Entropies of Lattice Defects*

H. B. HUNTINGTON, G. A. SHIRN,† AND E. S. WAJDA‡
Department of Physics, Rensselaer Polytechnic Institute, Troy, New York
(Received March 24, 1955)

The entropy of lattice defects are here considered in somewhat greater detail than previously. Calculations are out to estimate the intrinsic entropies associated with such lattice defects as interstitialcies and vacancies, and with the activated complexes of vacancy and ring diffusion. The present application is to fcc metals and a very simplified force model representing only the closed-shell, ion-core repulsions of copper has been used. The problem consists in evaluating the change of lattice vibrations with the introduction of each defect. The method is in a first approximation to consider vibrations localized around the defect as separate from elastic vibrations at appreciable distance from the defect. Only in the region far from the defect is the entropy contribution always positive. The surface condition of zero pressure introduces a term which reduces the effect of the "localized" vibrations. In general the intrinsic entropy is less positive for those configurations where there is crowding of the atoms than for those configurations where the atoms have greater free volume.

IN any calculation of the number of defects which may be expected in a crystal at thermal equilibrium at fixed temperature and pressure one proceeds by minimizing the Gibbs free energy of the system. By the general arguments of statistical mechanics one shows that the fraction of the lattice sites which may be associated with a particular lattice defect is given by $\exp(-\Delta G/RT)$, where ΔG is the change in the Gibbs potential of the crystal with the addition of one mole of defects. The wide use of this expression is a natural consequence of the very great importance of the role of defects in the most of the active fields of present solid state physics, such as optical properties, radiation damage, plastic properties and electrical properties. For those phenomena which depend not simply on the existence of the defects but upon their stepwise motion (diffusion, ionic conductivity, and some aspects of internal friction), the expression[1] for the rate change includes a similar factor, $\exp(-\Delta G^*/RT)$, where ΔG^* is the change in the Gibbs potential with the addition of one mole of defects leaving out of consideration the degree of freedom associated with the direction of motion. Hereafter the asterisk will be omitted here

* Supported by contract with the U. S. Atomic Energy Commission.
† Now at Sprague Electric Company, North Adams, Massachusetts.
‡ Now at Physics Department, Union College, Schnectady, New York.

[1] Glasstone, Laidler, and Eyring, *The Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York, 1941).

and the general features of both situations will be discussed in terms of a simple ΔG.

From the standard thermodynamic relation one shows

$$\Delta G=\Delta H-T\Delta S, \tag{1}$$

where ΔH is the change in enthalpy and ΔS is the change in entropy (exclusive of mixing entropy) resulting from the addition of a mole of defects. Since ΔH and ΔS appear in general quite temperature-independent,[2] it is customary to speak of a temperature-independent factor $\exp(\Delta S/R)$ and a temperature-dependent factor $\exp(-\Delta H/RT)$, where the nomenclature applies equally to finding the equilibrium number of defects or to the treatment of rate processes involving defects. Accordingly the evaluation of the ΔS in the temperature-independent factor can be approached from another fundamental thermodynamic formula to give

$$\Delta S=-(\partial\Delta G/\partial T)_p. \tag{2}$$

This problem has been treated several times in one form or another in the literature,[3] not always from consistent viewpoints. Individual aspects have received particular emphasis in special applications. On some points there has been duplication and confusion. For example the role of the thermal expansion in particular needs some clarification with respect to Eq. (2). One might at first think that ΔG would show a temperature dependence because of the change of volume-dependent atomic forces with the thermal expansion. Some care must be exercised here, however. For theoretical models which set up one term in the Helmholtz free energy, which represents the energy of the crystal at 0°K and is dependent on volume only, the said term can make no contribution to the entropy of the system since

$$\Delta S=-(\partial\Delta A/\partial T)_V, \tag{3}$$

where A is the Helmholtz potential. This point has been recently put forward by Vineyard and Dienes[4] but it is perhaps well to stress it here also.

The basic approach to this problem should be along the lines of statistical mechanics. The crystal and its defects can be fundamentally treated as a collection of reasonably harmonic oscillators whose vibrational contribution to A is given by

$$A_v=-kT\sum_i \ln\{1-\exp(-h\nu_i/kT)\}^{-1}, \tag{4}$$

where the ν_i are the frequencies of the crystal. Applying Eq. (3), one obtains

$$\Delta S=-k\sum_i\{\ln[1-\exp(-h\nu_i/kT)]^{-1}-(h\nu_i/kT)[\exp(h\nu_i/kT)-1]^{-1}\}, \tag{5}$$

or for $kT\gg h\nu_i$, which means T well above the Debye temperature,

$$\Delta S=-k\sum_i[\ln(h\nu_i/kT)+\text{terms of order unity}]. \tag{6}$$

If the original frequencies for the lattice are ν_{i0} and the final frequencies with defect are ν_{if}, then the complete change in entropy is

$$\Delta S=k\sum_i \ln(\nu_{i0}/\nu_{if}). \tag{7}$$

The problem of determining the specific entropy of a particular lattice defect (or activated complex) reduces to evaluating the change in the elastic spectrum that results from the introduction of the defect.

One procedure for accomplishing this would be to investigate the perturbations on the actual frequencies of the elastic waves of the crystal that the defect causes. A solution along these lines is being pursued by various investigators[5] and its successful completion will supply the only completely reliable treatment. The mathematical difficulties are, however, formidable. Our approach has been approximate but affords a program which should give some insight into the physical considerations that are involved. It proceeds from a consideration of localized vibrations as in the Einstein model. In this way one considers three regions: the atoms in the immediate neighborhood of the defect, the region over which the defect sets up an elastic stress field and finally the surface conditions and their influence.

For the immediate neighborhood of the defect an Einstein model is invoked whereby each atom near the defect is considered to vibrate individually as though in the field of its neighbors all held fixed. It is readily apparent that where the defect pushes atoms closer together than the equilibrium distance the frequencies rise and the ΔS is negative. Conversely, where the defect causes local expansion as in the case of a vacancy, one can expect the frequencies to drop and the ΔS to go positive. Again this statistical result finds its ready analog in the thermodynamic relation,

$$(\partial S/\partial V)_T=(\partial P/\partial T)_V>0. \tag{8}$$

The neglect of coupling between the vibrations inherent in the Einstein model is a serious limitation. Various devices have been used to compensate for its effect for the different defects and will be discussed later. In the region somewhat more removed from the defect, there is a radial strain falling off as r^{-2}. The strain field from this "center of pressure"[6] is pure shear with no dilatation. Since the strain is practically temperature-

[2] For a very complete experimental verification of the temperature independence of ΔS and ΔH in the diffusion of antimony in silver, see Sonder, Slifkin, and Tomizuka, Phys. Rev. **93**, 970 (1954).

[3] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (University Press, Oxford, 1948); J. Frenkel, *Kinetic Theory of Liquids* (Clarendon Press, Oxford, 1946), Chap. I; G. J. Dienes, Phys. Rev. **89**, 185 (1953); Y. Haven and J. H. van Santen, Phillips Research Repts. **7**, 474 (1952); Huntington, Shirn, and Wajda, Phys. Rev. **91**, 246(A) (1953).

[4] G. H. Vineyard and G. J. Dienes, Phys. Rev. **93**, 265 (1954).

[5] For example, K. F. Stripp and J. G. Kirkwood, J. Chem. Phys. **22**, 1579 (1954).

[6] A. E. H. Love, *Mathematical Theory of Elasticity* (Dover Publications, New York, 1944), fourth edition, Chap. XI.

independent, the stored elastic energy will have a temperature variation determined principally by the temperature variation of the elastic constants of the medium. Because the latter usually decrease with temperature, the contribution to the defect entropy from this region will be positive. This type of argument has been effectively used by Zener[7] and others in the past, sometimes with the implication that because of it all defect entropies should be positive. It would appear, however, that its validity is limited to those regions where the strains are elastic and the dilatation vanishes. Finally, the boundary conditions at the surface of the specimen have their influence on the defect entropy. In general the condition that the pressure vanishes requires that a constant dilatation be added to the strain pattern of the center of pressure with the sign of the dilatation the same as that of the local pressure, i.e., for an interstitial defect the dilatation is positive, conversely for a vacancy it is negative. The effect of a positive dilatation is to lower the lattice frequencies and again a contraction will raise them. In every case the effect of the surface boundary condition is to compensate in part for the entropy change associated with the immediate region of the defect.

FORCE MODEL AND ATOMIC VIBRATIONS

Application to the consideration of specific defects requires a force model which will give a reasonable characterization of atomic forces with a minimum of mathematical complexity. We have chosen to consider that, in the case of the metals at least, the cohesive forces are primarily volume-dependent and of long-range nature, relatively unaffected by the presence of defects. The repulsive forces are central in nature and arise from the repulsion of the closed ion shells. These vary rapidly with interatomic distance and for many materials make the dominant contribution to the elastic constants. As is quite usual, we shall represent the ion-core interaction by an exponential potential of the Born-Mayer[8] type,

$$U(r)=A\exp[-\rho(r-r_0)/r_0], \tag{9}$$

where A gives the magnitude of the interaction per ion pair at the equilibrium interatomic distance r_0 and ρ is the dimensionless constant whose magnitude is a measure of the rapidity with which the ion core interaction changes with distance. In this work we have taken ρ to be 13.6, a value which has been shown[9] to give a reasonable approximation of the elastic behavior of copper. Since we shall be here mainly concerned with ratios of frequencies, it will not be necessary to chose a value for A.

[7] C. Zener, J. Appl. Phys. **22**, 372 (1951); *Imperfections in Nearly Perfect Crystals* (John Wiley and Sons, Inc., New York, 1952), pp. 289–314.

[8] F. Seitz, *Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), Chap. II.

[9] C. Zener, Acta Cryst. **3**, 346 (1950); H. B. Huntington, Phys. Rev. **91**, 1092 (1953).

1. Einstein Model

Next, in calculating the elastic frequencies of the atoms near the defect on the basis of the Einstein model, one considers each atom to move in a potential minimum established by its interactions with each of its neighbors, presumed to be held fixed. The increase in energy acquired by the jth atom as it moves from its equilibrium position may be written

$$\Delta V_j=\sum_i \Delta V_{ij}(r_{ij}), \tag{10}$$

where i numbers the neighbors of the jth atom and ΔV_{ij} is the part of ΔV_j which is contributed by the ith atom,

$$\Delta V_i=U'(r)\delta_{11}+\delta_{11}{}^2U_0''(r)/2+\delta_\perp{}^2U'(r)/2r, \tag{11}$$

where δ_{11} and $\delta_\perp$ are the displacements of the jth atom, respectively parallel and perpendicular to the line of centers between i and j. For a collection of particles at stable equilibrium, the potential at the jth particle is

$$\Delta V_j=\tfrac{1}{2}\sum_i[U''\cos^2\theta_i+(U'/r)\sin^2\theta_i]\delta_j{}^2, \tag{12}$$

where δ_j is the displacement of the jth atom and θ_i is the angle that the displacement makes with the line of centers between atoms i and j. Our procedure is then to evaluate ΔV_j along the principal axes for each atom in units of $A\delta^2/r_0{}^2$ and compare the result with the value for the perfect face-centered lattice,

$$\Delta V_1=\tfrac{1}{2}\delta^2[4U''+8U'/r]_0=(2\rho^2-4\rho)A(\delta/r_0)^2. \tag{13}$$

This technique has been applied to the following defects: a single interstitial atom, first with the nearest neighbors at their normal lattice positions and later with these neighbors relaxed to new equilibrium position; a single vacancy, both unrelaxed and relaxed; a vacancy at the saddle point for migration, both unrelaxed and relaxed; and a ring mechanism at the saddle point. In a situation of high crowding, ΔV_j is large in the direction of compression because U'' is large and positive. This means increased frequency and lowered entropy. At right angles to the direction of compression, the frequencies are reduced by (U'/r) which is negative but an order of magnitude $1/\rho$ smaller than U''. For very severe uniaxial compressions it is possible for ΔV at right angles to compression axis to become small and go negative—corresponding to unstable equilibrium. Such a situation did occur for the unrelaxed nearest neighbors of the interstitial, but with relaxation stable equilibrium was restored and the geometric mean of all the frequencies increased. In general the U'/r term played a secondary role.

Because the frequencies vary as the square root of ΔV, the form of Eq. (7) useful for computation was

$$\Delta S=\tfrac{1}{2}k\sum_j \ln(\Delta V_1/\Delta V_j). \tag{7a}$$

The corresponding results are tabulated in the first two rows of Table I for unrelaxed and relaxed conditions respectively.

TABLE I. ΔS for various defects. (All values in units of k.)

Defect	Interstitial	Vacancy	Vacancy saddle point	Ring of four
Unrelaxed	—	1.85	−0.65	
Einstein model	— — — — — — — — — — — — — — — — — — — —			
Relaxed	−6.4	1.69	+0.87	−1.96
Coupling correction	0.2	0.09		
Elastic term	4.9	0.13	0.26	
Surface term	2.1	−0.44	−0.20	
Total	0.8	1.47	0.93	

For the interstitial configuration the frequencies of the interstitial and its six neighbors are analyzed. As a detailed example of the method values for ΔV for the interstitial atom, which is independent of direction, and for the neighbors for radial and tangential motion are shown for both the unrelaxed and relaxed cases in Table II. Because of the negative value of ΔV for the tangential motion of the neighbors, no value can be given for ΔS in the appropriate position in Table I. This example points up the importance of considering the relaxation, and in general it can be said that its effect is considerable, especially when crowding takes place. Here the relaxation consisted of a radial outward motion of the neighbors of about $0.15r_0$, determined as the position of minimum potential energy.

For the vacancy, the twelve nearest neighbors of the vacancy were considered and the effect of the vacancy on their radial and tangential frequencies taken into account. The relaxation consisted of an inward radial displacement of about $0.015r_0$ and reduces the "local" entropy from $1.85k$ to $1.69k$. In considering the saddle point for vacancy diffusion, one is interested in the activated complex which occurs when one of the atoms neighboring the vacancy moves halfway in to fill it. The moving atom passes through the center of a rectangle formed by four atoms at the corners at a distance of $(\sqrt{3}/2)r_0$ before relaxation. Because of the crowding the frequencies of these five atoms are increased, even after a relaxation outward of the corner atoms of about $0.06r_0$. The associated entropy decrease is $-1.10k$. There are, however, fourteen other atoms which are essentially in the environment of a neighbor of a vacancy and they contribute 7/6 the intrinsic entropy of a vacancy or $1.97k$. For the total entropy of the saddle point (always of course omitting from consideration the degree of freedom associated with the direction of diffusion), an increase of entropy of $0.87k$ results.

In the analysis of ring diffusion,[10] the radial frequencies of four partners in the ring and also their motion perpendicular to the plane of the ring were taken into account, also the motion of the eight atoms directly above and below the ring.

[10] C. Zener, Acta Cryst. 3, 346 (1950).

2. Effect of Coupling

Our procedure has been criticized because we have disregarded the normal modes of the lattice to treat the vibrations of the atoms as uncoupled. To gain some idea of the seriousness of this shortcoming, a normal mode solution was carried out for a seven-particle problem which simulated the configuration of an interstitial atom surrounded by six nearest neighbors. Four force constants were introduced to account for forces of the four following varities: (a) those between interstitial and neighbors, (b) those between neighbors, (c) those restraining the neighbors from radial motions away from equilibrium position, and (d) those restraining neighbors from tangential displacements. The labor involved was somewhat less than might at first be imagined. All but three of the frequencies were degenerate. The analysis required the solution of one cubic equation. The remaining frequencies were rational functions of the force constants. The product of all 21 frequencies was compared with the analogous product as obtained from the same set of constants using the usual approximation where coupling effects are neglected. Because of the relaxation the distances between neighbors were considerably extended and the coupling forces between them very weak. The principal effect came from the coupling of the motions of the interstitial and individual neighbors. Three sets of triply degenerate frequencies are involved, and their solution depends on a secular determinant of the form

$$\begin{vmatrix} a+c & -a & 0 \\ -a & 2a & -a \\ 0 & -a & a+c \end{vmatrix},$$

where a and c refer to the force constants mentioned earlier in this paragraph. The product of the roots is given by the value of the determinant. For the Einstein model (no coupling) the corresponding product is just the product of the diagonal elements. The ratio of the two is $c/(a+c)$, which is in our case 0.59. The corresponding entropy increase is $-(3k/2)\ln 0.59$, or $+0.81k$.

One must now compare this value with the increase in entropy for a complex of six atoms, stationed at the face centers of a cube to represent the normal lattice and subject to the force constants, b, c, and d. The normal frequencies for this aggregate are easily obtained. A comparison of the frequency product with and without coupling gives an increase of $0.38k$ in entropy. (In effect, coupling adds only nondiagonal

TABLE II. V's for the interstitial configuration in units of $A\beta^2/r_0^3$.

	V-interstitial	V-neighbor-radial	V-neighbor-tangential
Unrelaxed	7900	5290	−204
Relaxed	1050	1279	339

terms to the potential matrix. It always decreases the frequency product since it increases the frequency spread without changing the sum of the squares of the frequencies.) To the $0.38k$ entropy change must be added the entropy change of the average lattice atom with coupling, to take into account the seventh atom which was formerly the interstitial. It is shown in the Appendix that, on going from the Einstein to the Debye model, the average entropy per degree of freedom increases by $0.078k$. Adding three times this quantity to $0.38k$, one obtains $0.62k$ which is $0.19k$ less than the entropy increase with coupling for the complex with interstitial. It appears that the effect of coupling is small for the intrinsic entropy of an interstitial. In the third line of Table I, $0.2k$ is entered as an estimate of the magnitude of the correction involved.

A similar treatment for a vacancy and its twelve neighbors was not attempted because of the formidable complications associated with the large number of degrees of freedom. However, the two-dimensional analog, namely an hexagonal ring of atoms with and without a central atom, was investigated and the results proved illuminating. The radial and tangential displacements of the hexagon atoms were chosen as coordinates. Symmetry considerations indicate that the normal modes for this system, like that of six similar particles on a string, are characterized by representations of the form $\exp(isn/6)$, where n numbers the particles in rotation and s, running from 1 to 6, numbers the mode. For each value of s there are two modes depending on the phase relation between the radial and tangential motions. Modes with s equal to 1 and 5 are coupled together in two pairs of degenerate frequencies, and similarly for s equal to 2 and 4. The introduction of a central atom adds two more degrees of freedom. The coordinates of this atom represent a quantity with vector symmetry which couples only with the modes with s equal to 1 and 5. The solution of the resulting cubic equation gives three doubly degenerate modes. Evaluation of the frequency products gives for the hexagon with vacancy an entropy increase of $0.46k$ with coupling, and for the filled hexagon (less the entropy change for one normal atom with coupling) $0.37k$, or a net change of $0.09k$ for the vacancy in the hexagon with coupling. It should be pointed out that only one force constant is involved here and that the result is independent of its magnitude. Again the effect of coupling is nearly negligible on the defect entropy. As an indication of its size, $0.09k$ is entered for the vacancy on the third line of Table I.

In their calculation on the effect of vacancies on lattice entropy, Stripp and Kirkwood[6] give a first order contribution for the isolated vacancy without relaxation of $1.5k$ and a second-order contribution of $0.225k$. For comparison with their value of $1.73k$, the Einstein model with central forces gives

$$\Delta S = -zk/2\ln(1-3/z) = 1.68k$$

for $z=12$. Adding on the estimated correction of $0.09k$, one obtains $1.77k$ which is certainly fortuitously close. The discrepancy between $1.68k$ and the value of $1.85k$ listed in Table I arises because our model tacitly assumes, in addition to central forces, a volume-dependent, cohesive force to maintain stability.

3. Elastic Energy Contribution

To calculate the contribution to the defect entropy arising from regions somewhat removed from the defect, one can resort to the thermodynamic method of Zener[7] and apply Eq. (2) using the stored energy for ΔG. This is permissible in this region because elastic theory is valid here and the elastic solution for a center of pressure gives no dilatation. The general form for the displacement from such a center is[8]

$$\xi_i = x_i(A + B/r^3), \tag{14}$$

where the coefficient A gives a uniform dilatation and the B term gives the strain pattern from the center of pressure in an infinite medium. From the displacements one can obtain the strain and the stresses. The resulting expression for the elastic energy density follows:

$$\text{Energy density} = \tfrac{1}{2}[(3\lambda+2\mu)A^2 + (2B)^2\mu/r^6], \tag{15}$$

where λ and μ are the Lamé elastic constants. The condition at the surface determines A. For zero pressure at great radius R, one obtains

$$A = 4B\mu/(3\lambda+2\mu)R^3, \tag{16}$$

which shows that the A term in Eq. (15) is negligible compared with the B term. The total elastic energy outside a sphere of radius r_1 is then approximately

$$\Delta G = 8\pi\mu B^2/r_1^3, \tag{17}$$

and the contribution to ΔS is

$$\Delta S_{\text{elas}} = (-8\pi B^2/r_1^3)(d\mu/dT). \tag{18}$$

This expression has been evaluated for copper, taking for $d\mu/dT$ -0.38×10^9 dynes/cm²-deg as an average between -0.45 and -0.28×10^9 dynes/cm²-deg, respectively the derivatives with respect to temperature of the elastic constants c_{44} and $\frac{1}{2}(c_{11}-c_{12})$ for copper.[11] The constant B is determined from our assumed relaxation of defect neighbors and for the interstitial is taken to have the value $0.025d^3$, for the vacancy $-0.0053d^3$, where d is the lattice constant, $3.61A$. The radius r_1 is chosen so that $4\pi r_1^3/3$ is the volume containing the same number of atoms as were considered in the normal mode treatment, seven for the interstitial case and twelve for the vacancy. The resulting value for B^2/r_1^3 was $15.0\times10^{-4}d^3$ for the interstitial and $3.9\times10^{-5}d^3$ for the vacancy. The values for ΔS_{elas}

[11] These experimental values for the variation of the elastic constants of copper with temperature were supplied by Dr. C. S. Smith by private communication.

are given in the fourth row of Table I as $4.9k$ for the interstitialcy and $0.13k$ for the vacancy.

Configurations not possessing spherical symmetry cannot be treated as simply. Because of the importance of the vacancy saddle point in the theory of diffusion we have made an effort to handle this case by introducing a third term in the elastic potential,[8] so that

$$\chi=\tfrac{1}{2}Ar^2-B/r+C(3\cos^2\theta-1)/2r^3. \quad (19)$$

The gradient of χ gives the displacement vector but now, in addition to the terms appearing in Eq. (14), there is an additional term in the radial component varying with angle and a nonradial component which arise from the term in the second spherical harmonic. The constants are chosen to give a radial displacement of $-0.015r_0$ for $r=1.5r_0$, $\theta=0$ and of $0.06r_0$ for $r=0.87r_0$, $\theta=\pi/2$. (The direction of the θ axis coincides with the direction of the diffusion jump.) One obtains

$$B=-0.0024r_0^3, \quad 3C=0.072r_0^5. \quad (20)$$

The contribution to the energy density to be added to Eq. (15) is

$$\tfrac{1}{2}(3C/r^3)^2(45\cos^4\theta-10\cos^2\theta+13)\mu. \quad (21)$$

The additional contribution to Eq. (17) is

$$\Delta G(C)=16.7(3C)^2\mu/r_1^7, \quad (22)$$

where r_1 is the spherical radius needed to enclose 19 atoms, and to Eq. (18) add

$$\Delta S(C)=-16.7(3C)^2/r_1^7(d\mu/dT). \quad (23)$$

The evaluation of Eqs. (18) and (23) for the appropriate choices of B, C, and r_1 gives respectively $0.002k$ and $0.262k$, or a total of $0.26k$ for the entropy change of the vacancy saddle point arising from the stored elastic energy outside r_1.

4. Surface Effect

As has already been pointed out in the preceding section the elastic solution for a center of pressure contains a term corresponding to a uniform dilatation which is introduced to fit the boundary conditions at the surface of zero pressure. Though this term makes a negligible contribution to the elastic energy of the deformation it makes a considerable contribution to the defect entropy because the elastic constants, and hence the vibration frequencies, are decreasing functions of volume. For the long wave vibrations the change in the frequencies, ν_i, are

$$\Delta\nu_i/\nu_i=\tfrac{1}{2}(\Delta M_i/M_i+\Delta V/V), \quad (24)$$

where M_i are the appropriate elastic moduli. For copper the effect of pressure on the elastic constants has been measured by Lazarus[12] and found to be linear out to 10^4 bars. At this maximum pressure the fractional changes in the elastic constants were 0.028 for c_{11}, 0.03_0 for c_{12}, and 0.01 for c_{44} with a corresponding fractional change of volume of -0.007. Taking 0.023 as an average fractional change of elastic modulus at 10^4 bars, one obtains

$$\Delta\nu_i/\nu_i=\tfrac{1}{2}[(0.023/0.007)-1]\Delta V/V. \quad (24a)$$

With Eqs. (7) and (16) and under the rough assumption that all frequencies are similarly affected by the dilatation $3A$, the entropy change is

$$\Delta S_{surf}=+3.4(3N)Ak, \quad (25)$$

where N is the total number of atoms in the lattice, $16R_1^3/3d^3$. Evaluation of S_{surf} gives $2.05k$ for the interstitial and $-0.44k$ for the vacancy. For the vacancy saddle point configuration only the B-term affects A (16) and $\Delta S=-0.20k$, where we have assumed surface equilibrium for the saddle-point configuration.[13]

The surface contribution to the entropy could also be calculated[14] directly from the macroscopic relation,

$$(\partial S/\partial V)_T=(\partial P/\partial T)_v=K\beta, \quad (26)$$

where K is the bulk modulus and β is the thermal coefficient of expansion. Evaluated on this basis, the surface terms would be larger by about 70 percent than the values which have been obtained here with the microscopic model, used throughout for consistency.

DISCUSSION

In the last line of Table I are given the totals for ΔS in three important configurations. One must warn against the literal acceptance of these figures at their face value. The calculations in back of the numbers is uncertain (1) because a greatly oversimplified model has been used to represent the atomic forces and (2) because in the calculations themselves some arbitrary approximations were necessarily introduced at such points as the separation of frequencies into "local" and "elastic" modes, the treatment of the effect of coupling, the fitting of the elastic solution, and others. From a critical standpoint the procedure here is not to be considered as a handbook example on the quantitative calculation of defect entropies, but rather as a qualitative display of the way various considerations might enter into a more precise evaluation. From this standpoint several interesting conclusions can be drawn:

(1) The evaluation of defect entropies from a fundamental standpoint devolves into an analysis of the change of the elastic spectrum of the crystal with the introduction of the defect.

(2) Defects which bring about a local crowding of the lattice appear to have a lower specific entropy than

[12] David Lazarus, Phys. Rev. **76**, 553 (1949).

[13] Because of the short time the diffusing atom spends at the saddle point configuration it is perhaps quite unrealistic to consider any change in "surface term" during the jump.

[14] We are indebted to Dr. J. D. Eshelby for pointing out this alternate procedure.

defects which cause expansion. Here the effect of the "local" term is predominant, though it is partially compensated for by the effect of the dilatation introduced to satisfy the condition at the surface. The ΔS_{surf} has always the opposite sign from the "local" term and compensates for about $\frac{1}{4}$ of it.

(3) Though it appears that most of the defects may have positive entropies, the role of the temperature dependence of the elastic constants in causing this effect appears to have been overemphasized in some quarters. Our calculations show that the size of this term as applied only to the elastic region is often rather small. Since it varies as the square of the displacement it becomes most important in situations involving large distortions, as for an interstitialcy.

(4) The errors involved in using an Einstein model for calculating frequency changes appear to be small as indicated by efforts to take coupling into account to the first order.

(5) The comparisons of these estimates with experimental findings on defect entropies, particularly as they come out of diffusion studies, is not altogether satisfactory. For a vacancy mechanism the temperature-independent factor in the diffusion constant should depend exponentially on the entropy of the vacancy saddle-point configuration. In general, experiments show the entropy to be positive but somewhat larger than the $0.9k$ obtained here. Also, the reason for the fairly consistent correlation between the sizes of activation entropy and activation energy, as experimentally observed, is not easily apparent in terms of the present model. This correlation has been displayed in various forms by different investigators, beginning with the well-known Langmuir-Dushman formula and including the semiempirical relation proposed by Dienes[15] and the recent theory of Zener.[7] The experiments indicate that the largest positive entropies are found for those materials with the largest activation energies. Of the terms in Table I that make up ΔS for the vacancy saddlepoint, the relatively small "elastic" term for the region surrounding the defect should support such a correlation between activation entropy and activation energy. What the effect of the term depending on the changes in the "local vibrations" should be is not so easy to see. We have stressed the importance of the change in microscopic density for this quantity. Translated loosely into macroscopic thermodynamic terms, the effect of such density changes can be estimated from Eq. (26). The changes in "local vibrations" therefore should apparently depend in part on the relation (26) and in part on the temperature variation of elastic constants. In each contribution the magnitude of the crystal forces and the effect of thermal expansion play their roles so that the aforementioned correlation between activation energy and entropy may be also understandable in terms of this theory (as well as in terms of a theory which considers only the temperature variation of elastic constants).

[15] G. J. Dienes, J. Appl. Phys. **21**, 1189 (1950).

APPENDIX

The change in entropy per degree of freedom in going from the Einstein to the Debye vibration model can be estimated by assuming for the Debye model that the density of oscillators per frequency range $d\nu$ is given by $K\nu^2$, where K is normalized so that

$$\int_0^{\nu_m} K\nu^2 d\nu = K\nu_m^3/3 = n, \tag{A1}$$

where n is the total number of oscillators. On the other hand, the Einstein spectrum is a delta function centered at some frequency ν_E, so that the sum of the frequencies squared is unchanged,

$$n\nu_E^2 = K\int_0^{\nu_m} \nu^4 d\nu = 3n\nu_m^2/5. \tag{A2}$$

It follows that ν_E/ν_m is 0.77. To find the entropy increase one evaluates the average of the logarithm of the frequency ratio,

$$S = -(kK/n)\int_0^{\nu_m} \ln(\nu/\nu_E)\nu^2 d\nu = 0.078k. \tag{A3}$$

The analogous quantity for a two-dimensional distribution is $0.153k$. In one dimension there is a logarithmic singularity.

Electronic States in a Perturbed Periodic Lattice*

PAULA FEUER
Purdue University, Lafayette, Indiana
(Received April 29, 1955)

Solutions of the one-dimensional Schrödinger equation for a periodic potential modified by a perturbing potential are expressed as $\psi(x)=\sum_k \phi(x_k)\, a(x-x_k)$. The function $a(x-x_k)$ is a Wannier function localized around the kth atom of the crystal and associated with a particular permitted band; the coefficients ϕ satisfy a differential equation of infinite order. W.B.K. solutions of the equation for ϕ are obtained. In the permitted band, ϕ is oscillatory; in the neighboring forbidden bands ϕ decreases exponentially with distance from a band edge. Various types of perturbing potentials are considered and quantum conditions are derived for impurity states and for states in which the permitted band is crossed.

I. INTRODUCTION

CONSIDER the solution of the one-dimensional Schrödinger equation for a periodic potential $V_p(x)$ modified by a perturbing potential $V(x)$,

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2}+[V_p(x)+V(x)]\psi=E\psi. \tag{1}$$

It is well known that when $V(x)=0$, two linearly independent solutions of Eq. (1) have the form

$$\psi_0(\pm p;x)=e^{\pm ipx/\hbar}P_0(\pm p;x), \tag{2}$$

corresponding to the energy $E_0(p)$, where the P_0's are functions with the periodicity of the lattice. In a permitted energy band, p is real; in a forbidden band p is imaginary.

The nature of the solutions of Eq. (1) has been considered in detail by James.[1] His method consists of joining together solutions valid for a single period of the unperturbed potential. The method however is not applicable to perturbations in which a permitted band is crossed.

Slater[2] has shown that the solutions of Eq. (1), for the case in which the energy lies always near a particular permitted band, can be expressed in the form

$$\psi(x)=\sum_k \phi(x_k)a(x-x_k). \tag{3}$$

The $a(x-x_k)$ are Wannier functions associated with the particular permitted band and localized around the kth atom of the crystal; they are defined by

$$a(x-x_k)=\sum_p e^{ipx_k/\hbar}\psi_0(p;x), \tag{4}$$

the sum being taken over all values of the effective momentum p in the permitted band. The $\phi(x_k)$ are coefficients which give essentially the amplitude and phase of the wave function in the neighborhood of the kth atom.

There are two alternative procedures for determining these coefficients. Slater has derived a set of difference equations for the $\phi(x_k)$, and these have been solved for several special types of perturbing potentials.[3-5] On the other hand it has been shown[2] that the ϕ, considered as a continuous function of x, must satisfy the differential equation

$$E_{op}\phi(x)+V(x)\phi(x)=E\phi(x). \tag{5}$$

E_{op} is a differential operator of infinite order, defined in the following way. In the unperturbed case, $E_0(p)$ can be considered as a periodic function of the effective momentum p across a permitted band. E_{op} is then obtained by replacing p by $-i\hbar d/dx$ in the expression for $E_0(p)$; that is

$$E_{op}=E_0(-i\hbar d/dx). \tag{6}$$

Solutions of Eq. (5) have been considered only for the case in which the energy lies always close to a band edge. In this case, E_{op} can be approximated by a second-order operator, Eq. (5) becomes the "effective mass equation," which has the form of the ordinary Schrödinger equation, and conventional methods can be used to solve it.

It will now be shown how W.B.K. solutions of Eq. (5) can be obtained. Various types of perturbing potentials will then be considered and quantum conditions will be derived for impurity states and for states in which a permitted band is crossed.

II. W.B.K. SOLUTIONS OF THE EQUATION FOR $\phi(x)$

Figure 1 shows how $E_0(p)$ varies with p as the permitted band is crossed. E_{c1} and E_{c2} are, respectively, the lower and upper band edge energies. If $E_0(p)$ is considered as a periodic function of p, it can be written in the form

$$E_0(p)=E_0(-p)=\sum_n b_n\cos(pna/\hbar), \tag{7}$$

where the b_n are constants which depend on the specific form of the periodic potential $V_p(x)$. $E_0(p)$ can then also be written in the form

$$E_0(p)=\sum_n c_n p^n, \tag{8}$$

where again the c_n depend on the form of $V_p(x)$. By

* This work was supported in part by the Signal Corps.
[1] H. M. James, Phys. Rev. **76**, 1611 (1949).
[2] J. C. Slater, Phys. Rev. **76**, 1592 (1949).
[3] P. Feuer, Phys. Rev. **88**, 92 (1952).
[4] G. F. Koster and J. C. Slater, Phys. Rev. **95**, 1167 (1954).
[5] G. F. Koster, Phys. Rev. **95**, 1436 (1954).

using Eqs. (6) and (8), Eq. (5) becomes

$$\sum_n c_n(-i\hbar d/dx)^n\phi(x)+V(x)\phi(x)=E\phi(x). \quad (9)$$

Before considering solutions of Eq. (9), it is convenient to define a function $P(E-V)$, which depends on the perturbing potential V, by

$$E-V=\sum_n c_n P^n. \quad (10)$$

It should be noted that P is a multiple-valued function of $E-V$ for, if a function P satisfies Eq. (10), so does $\pm P+(2m\pi\hbar/a)$, $m=0, \pm1, \pm2, \cdots$.

The perturbing potential V varies with x. For a given value of x, the perturbing potential has a value $V(x)$ and to this value of $V(x)$ there correspondes a P which satisfies Eq. (10). Thus we can define a (multiple-valued) function $P(x)$ such that

$$P(x)=P(E-V(x)).$$

$P(x)$ then satisfies the relation

$$E-V(x)=\sum_n c_n P^n(x). \quad (11)$$

Now let us follow the usual procedure for obtaining a W.B.K. solution and assume that the solutions of Eq. (9) are of the form

$$\phi(x)=e^{iS(x)/\hbar}; \quad S(x)=S_0(x)+(\hbar/i)S_1(x)+\cdots. \quad (12)$$

Then

$$(-i\hbar)^n\frac{d^n}{dx^n}\phi(x)=(-i\hbar)^n e^{iS(x)/\hbar}\left[\left(\frac{i}{\hbar}\right)^n\left(\frac{dS}{dx}\right)^n+\left(\frac{i}{\hbar}\right)^{n-1}\times\frac{n(n-1)}{2}\left(\frac{dS}{dx}\right)^{n-2}\frac{d^2S}{dx^2}+\cdots\right]$$

$$=e^{iS(x)/\hbar}\left[\left(\frac{dS_0}{dx}\right)^n+\hbar\left\{n\left(\frac{dS_0}{dx}\right)^{n-1}\frac{dS_1}{dx}+\frac{n(n-1)}{2}\left(\frac{dS_0}{dx}\right)^{n-2}\frac{d^2S_0}{dx^2}+\cdots\right\}\right].$$

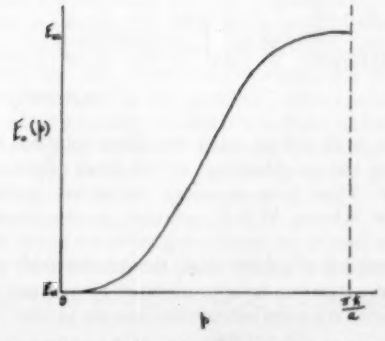


FIG. 1. Dependence of E_0 on the effective momentum p across a permitted band.

Equation (9) then becomes

$$\left[\sum_n c_n\left[\left(\frac{dS_0}{dx}\right)^n+\hbar\left\{n\left(\frac{dS_0}{dx}\right)^{n-1}\frac{dS_1}{dx}+\frac{n(n-1)}{2}\times\left(\frac{dS_0}{dx}\right)^{n-2}\frac{d^2S_0}{dx^2}\right\}\right]+\cdots-\{E-V(x)\}\right]e^{iS(x)/\hbar}=0. \quad (13)$$

Setting the coefficient of $\hbar^0$ equal to zero, we obtain

$$\sum_n c_n(dS_0/dx)^n-[E-V(x)]=0.$$

The solution of this, as indicated by Eq. (11) can be written

$$dS_0/dx=P(x). \quad (14)$$

Thus,

$$S_0(x)=\int P(x)dx+C. \quad (15)$$

Similarly, setting the coefficient of $\hbar$ in Eq. (13) equal to zero and using Eq. (14) we obtain an equation for $S_1(x)$,

$$\frac{dS_1}{dx}\sum_n c_n nP^{n-1}+\frac{dP}{dx}\sum_n\frac{n(n-1)}{2}c_nP^{n-2}=0. \quad (16)$$

The sums in Eq. (16) are easily evaluated. If Eq. (10) is differentiated with respect to P we obtain

$$-dV/dP=\sum_n c_n nP^{n-1}, \quad (17)$$

and differentiating Eq. (17) with respect to P we obtain

$$-d^2V/dP^2=\sum_n c_n n(n-1)P^{n-2}. \quad (18)$$

Thus, using Eqs. (17) and (18), Eq. (16) becomes

$$\frac{dS_1}{dx}\frac{dV}{dP}+\tfrac{1}{2}\frac{dP}{dx}\frac{d^2V}{dP^2}=0,$$

the solution of which is

$$S_1=-\tfrac{1}{2}\ln(dV/dP). \quad (19)$$

Then, to the approximation considered, using Eqs. (12), (15), and (19), $\phi(x)$ has the form

$$\phi(x)=C\left(\frac{dV}{dP}\right)^{-1}\exp\left(\frac{i}{\hbar}\int P(x)dx\right). \quad (20)$$

One can expect this to be a good approximate solution except in regions near the zeros of dV/dP.

A. Solutions in the Permitted Band

Figure 2 illustrates how a permitted band might be distorted by a perturbing potential; we first consider the case such that the energy E lies everywhere in the

FIG. 2. Energy band distorted by a perturbing potential.

permitted band. If $p(x)$ is a function that satisfies Eq. (11) and such that its magnitude lies between 0 and $\pi\hbar/a$, the general form of the function that satisfies Eq. (11) is

$$P(x)=\pm p(x)+2m\pi\hbar/a, \quad m=0,\pm1,\cdots. \tag{21}$$

Thus, from Eqs. (20) and (21),

$$\phi(x)=C\left\{\frac{dV}{dp}\right\}^{-\frac{1}{2}}\exp\left(\pm\frac{i}{\hbar}\int p(x)dx\right)e^{2\pi mxi/a}. \tag{22}$$

The function dV/dp, when written in the form

$$dV/dp=-d\{E-V\}/dp,$$

lends itself more easily to physical interpretation. Bloch has shown that, in the unperturbed crystal, the electron can be represented by a wave packet; if the wave packet is constructed from Bloch waves with effective momenta in the neighborhood of p, the velocity of the wave packet is given by $dE_0(p)/dp$. Thus, in the perturbed crystal,

$$d[E-V]/dp=v(x) \tag{23}$$

represents the local value of this wave packet velocity, and $\phi(x)$ can be written in the form

$$\phi(x)=\frac{C}{[v(x)]^{\frac{1}{2}}}\exp\left(\pm\frac{i}{\hbar}\int p(x)dx+\frac{2\pi mxi}{a}\right). \tag{24}$$

The infinite number of solutions indicated by Eq. (24) is to be expected since Eq. (5) is an equation of infinite order. However, at the lattice points, where we are interested in the functions $\phi(x)$, the solution of Eq. (5) can be written as a linear combination of the two functions,

$$\phi(x)=\frac{C}{[v(x)]^{\frac{1}{2}}}\exp\left(\pm\frac{i}{\hbar}\int p(x)dx\right). \tag{25}$$

When $V(x)=0$, $p(x)$, and $v(x)$ are constant and the expression for ϕ at the lattice points has the general form

$$\phi(x_k)=c_1e^{ipx_k/\hbar}+c_2e^{-ipx_k/\hbar}, \tag{26}$$

which when inserted into Eq. (3) yields

$$\psi(x)=c_1\psi_0(p;x)+c_2\psi_0(-p;x). \tag{27}$$

As one would expect, this is a linear combination of the two linearly independent unperturbed wave functions.

It is interesting to compare Eq. (25) with the W.B.K. solutions of the ordinary Schrödinger equation for a perturbing potential $V(x)$.[6] In that case, the wave functions have the same form as Eq. (25); however $p(x)=\{2m[E-V(x)]\}^{\frac{1}{2}}$ and $v(x)=p(x)/m$. Thus, replacing the second-order operator of the Schrödinger equation by the infinite-order operator of Eq. (5), replaces the classical momentum by the effective momentum, and the classical velocity by the wave packet velocity.

At the band edges, $v(x)=0$. Thus, the W.B.K. solutions break down in the neighborhood of the band edges. (This is analogous to the breaking down of the W.B.K. solutions of the ordinary Schrödinger equation in the neighborhood of a classical turning point.) In the neighborhood of a band edge, however, Eq. (5) can be approximated by a second-order differential equation whose solutions will be discussed in Sec. II. C.

B. Solutions in the Forbidden Bands

In the forbidden band adjacent to the $p=0$ band edge where $E-V(x)<E_{e1}$, the lower band edge energy, Eq. (11) will continue to hold if we allow $P(x)$ to become imaginary. Again we can write

$$P(x)=\pm p(x)+2m\pi\hbar/a, \tag{28}$$

but here $p(x)$ is pure imaginary; $|p(x)|$ is zero at the band edge and increases as $E-V(x)$ decreases. The proof of the W.B.K. solution then follows as before; the solutions have the general form of Eq. (24) and can also be written

$$\phi(x)=\frac{C}{(|v(x)|)^{\frac{1}{2}}}\exp\left(\pm\frac{1}{\hbar}\int|p(x)|dx\right)e^{2\pi mxi/a}. \tag{29}$$

The function $v(x)$, given by Eq. (23), is pure imaginary.

In the forbidden band adjacent to the $p=\pi\hbar/a$ band edge where $E-V(x)>E_{e2}$, Eq. (11) continues to hold for $P(x)$ of the form of Eq. (28) but with $p(x)$ such that $[\pi\hbar/a-p(x)]$ is pure imaginary and such that $|\pi\hbar/a-p(x)|$ is zero at the band edges and increases as $E-V(x)$ increases. The function $v(x)$ is again pure imaginary and $\phi(x)$ has the general form

$$\phi(x)=\frac{Ce^{ixx/a}}{(|v(x)|)^{\frac{1}{2}}}\exp\left(\pm\frac{1}{\hbar}\int\left|\frac{\pi\hbar}{a}-p(x)\right|dx\right)$$
$$\times e^{2\pi mxi/a}. \tag{30}$$

Again it should be noted that these solutions break down in the neighborhood of the band edges where $v(x)=0$. Thus it is necessary to derive connection formulas relating W.B.K. solutions in the permitted band to those in the neighboring forbidden bands. In the neighborhood of a band edge, the infinite-order differential equation can be approximated by a second-order differential equation whose solutions are known. These

[6] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), pp. 178–187.

equations will be used to obtain the desired connection formulas.

C. Solutions in the Neighborhood of the Band Edges

1. *Solutions near the $p=0$ Band Edge*

Near the $p=0$ band edge, $E_0(p)$ is approximately

$$E_0(p)=E_{c1}+p^2/2m_1^*, \tag{31}$$

where E_{c1} is the band edge energy and m_1^* is the effective mass associated with this band edge. The infinite-order operator can be approximated by

$$E_{\text{op}}=E_{c1}-\frac{\hbar^2}{2m_1^*}\frac{d^2}{dx^2}. \tag{32}$$

The function $p(x)$ is determined from

$$E-V(x)=E_{c1}+p^2(x)/2m_1^*, \tag{33}$$

and

$$v(x)=d\{E-V\}/dp=p(x)/m_1^*. \tag{34}$$

Moreover, the infinite-order differential equation can be approximated by [Eqs. (5) and (32)] the so-called "effective mass equation":

$$E_{c1}\phi(x)-\frac{\hbar^2}{2m_1^*}\frac{d^2\phi(x)}{dx^2}+V(x)\phi(x)=E\phi(x),$$

or

$$\frac{d^2\phi(x)}{dx^2}+\frac{p^2(x)}{\hbar^2}\phi(x)=0. \tag{35}$$

One can expect this equation to be a good approximation provided that the potential does not vary too rapidly near the band edge.

Equation (35) has the form of the ordinary Schrödinger equation and its W.B.K. solutions are well known.[6] They can be written

$$\phi(x)=\frac{C}{[v(x)]^{\frac{1}{2}}}\exp\left(\pm\frac{i}{\hbar}\int p(x)dx\right), \tag{36}$$

where $v(x)=p(x)/m_1^*$. In the permitted band $p(x)$ is real and the solutions, Eq. (36), join smoothly to solutions of the form of Eq. (24). In the adjacent forbidden band, $p(x)$ is pure imaginary and the solutions have the form of Eq. (29). Moreover the connection formulas relating solutions in the permitted band (classical region) where p is real to those in the forbidden band (nonclassical region) where p is imaginary are known.[6] These will be used in Secs. III and IV to derive quantum conditions for energy states corresponding to various types of perturbing potentials.

2. *Solutions near the $p=\pi\hbar/a$ Band Edge*

Near the $p=\pi\hbar/a$ band edge, $E_0(p)$ is approximately

$$E_0(p)=E_{c2}+\frac{(\pi\hbar/a-p)^2}{2m_2^*}, \tag{37}$$

where m_2^* is the (negative) effective mass associated with this band edge. The infinite-order operator can be approximated by

$$E_{\text{op}}=E_{c2}+\frac{1}{2m_2^*}\left\{(-i\hbar)^2\frac{d^2}{dx^2}-\frac{2\pi\hbar}{a}\left(-i\hbar\frac{d}{dx}\right)+\frac{\pi^2\hbar^2}{a^2}\right\}. \tag{38}$$

The function $p(x)$ is determined from

$$E-V(x)=E_{c2}+\frac{[\pi\hbar/a-p(x)]^2}{2m_2^*}, \tag{39}$$

and

$$v(x)=-\frac{[\pi\hbar/a-p(x)]}{m_2^*}. \tag{40}$$

The equation which approximates the infinite-order equation is then

$$-\frac{\hbar^2}{2m_2^*}\frac{d^2\phi}{dx^2}+\frac{i\pi\hbar^2}{am_2^*}\frac{d\phi}{dx}+\frac{\pi^2\hbar^2}{2m_2^*a^2}\phi-[E-E_{c2}-V(x)]\phi=0. \tag{41}$$

It is convenient to define a function $\chi(x)$ by

$$\phi(x)=e^{i\pi x/a}\chi(x). \tag{42}$$

Substituting Eq. (42) into Eq. (41), we obtain an equation for $\chi(x)$ which is

$$\frac{d^2\chi(x)}{dx^2}+\frac{[\pi\hbar/a-p(x)]^2}{\hbar^2}\chi(x)=0. \tag{43}$$

The W.B.K. solutions of this equation are

$$\chi(x)=\frac{C}{[v(x)]^{\frac{1}{2}}}\exp\left[\pm\frac{i}{\hbar}\int\left(\frac{\pi\hbar}{a}-p(x)\right)dx\right] \tag{44}$$

and the corresponding solutions for $\phi(x)$ are

$$\phi(x)=\frac{C}{[v(x)]^{\frac{1}{2}}}\exp\left(\frac{i}{\hbar}\int p(x)dx\right) \tag{45a}$$

and

$$\phi(x)=\frac{C}{[v(x)]^{\frac{1}{2}}}e^{i\pi x/a}\exp\left(-\frac{i}{\hbar}\int p(x)dx\right), \tag{45b}$$

where $v(x)$ is given by Eq. (37). Again as we would expect, these solutions have the form of Eq. (24) in the permitted band and the form of Eq. (30) in the forbidden band. Moreover, since Eq. (43) is of the Schrödinger type, the connection formulas relating the χ (and hence the ϕ) in the permitted and forbidden regions are known. These will be used in the following section to discuss impurity states.

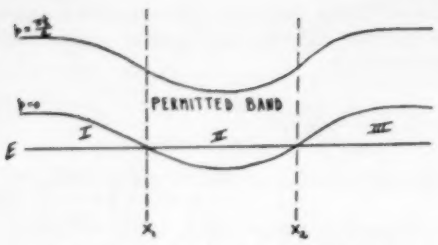


FIG. 3. Impurity state near a $p=0$ band edge.

III. IMPURITY STATES

A. Impurity States Near the $p=0$ Band Edge

Suppose the perturbing potential is of the type indicated in Fig. 3. In the neighborhood of the $p=0$ band edge, $\phi(x)$ has the form of Eq. (36). The connection formula for a solution that is well-behaved in region I is given by[6]

$$\frac{1}{2}\frac{1}{(|v(x)|)^{\frac{1}{2}}}\exp\left(-\frac{1}{\hbar}\int_x^{x_1}|p(x)|dx\right)\rightarrow \phi_{\mathrm{I}}(x)$$

$$\frac{1}{[v(x)]^{\frac{1}{2}}}\cos\left[\int_{x_1}^{x}\frac{p(x)}{\hbar}dx-\frac{\pi}{4}\right], \quad (46)$$

$$\phi_{\mathrm{II}}(x)$$

where $v(x)=p(x)/m_1^*$. $\phi_{\mathrm{II}}(x)$ joins smoothly to a linear combination of functions of the form of Eq. (24) and $\phi_{\mathrm{I}}(x)$ joins smoothly to a function of the form of Eq. (29), with $v(x)$ and $p(x)$ assuming the more general forms given by Eqs. (23) and (11), respectively.

The connection formula for a solution that is well-behaved in region III is given by

$$\frac{1}{2}\frac{1}{[|v(x)|]^{\frac{1}{2}}}\exp\left(-\frac{1}{\hbar}\int_{x_2}^{x}|p(x)|dx\right)\rightarrow \phi_{\mathrm{III}}(x)$$

$$\frac{1}{[v(x)]^{\frac{1}{2}}}\cos\left[\int_{x}^{x_2}\frac{p(x)dx}{\hbar}-\frac{\pi}{4}\right]. \quad (47)$$

$$\phi_{\mathrm{II}}(x)$$

The two expressions for $\phi_{\mathrm{II}}(x)$ given in Eqs. (46) and (47) can be made to join smoothly if

$$\int_{x_1}^{x_2}\frac{p(x)dx}{\hbar}=(m+\tfrac{1}{2})\pi, \quad (48)$$

where m is an integer. This is just the Bohr quantum condition that one might expect, except that $p(x)$ is not the classical momentum but is the local effective momentum that varies in a much more complicated way [Eq. (11)] with $V(x)$.

B. Impurity States Near a $p=\pi\hbar/a$ Band Edge

Suppose the perturbing potential is of the form indicated in Fig. 4. In the neighborhood of the $p=\pi\hbar/a$ band edge $\phi(x)$ has the form of Eq. (42) with χ given by Eq. (44). The connection formula for the solution χ which is well-behaved in region I is given by

$$\frac{1}{2}\frac{1}{(|v(x)|)^{\frac{1}{2}}}\exp\left(\frac{1}{\hbar}\int_x^{x_1}\left|\frac{\pi\hbar}{a}-p(x)\right|dx\right)\rightarrow$$

$$\frac{1}{[v(x)]^{\frac{1}{2}}}\cos\left[\frac{1}{\hbar}\int_{x_1}^{x}\left(\frac{\pi\hbar}{a}-p(x)\right)dx-\frac{\pi}{4}\right],$$

where $v(x)=-[\pi\hbar/a-p(x)]/m_2^*$. The corresponding connection formula for the $\phi(x)$ is

$$\frac{1}{2}\frac{e^{i\pi x/a}}{(|v(x)|)^{\frac{1}{2}}}\exp\left(\frac{1}{\hbar}\int_x^{x_1}\left|\frac{\pi\hbar}{a}-p(x)\right|dx\right)\rightarrow \phi_{\mathrm{I}}(x)$$

$$\frac{e^{i\pi x/a}}{[v(x)]^{\frac{1}{2}}}\cos\left[\frac{1}{\hbar}\int_{x_1}^{x}\left(\frac{\pi\hbar}{a}-p(x)\right)dx-\frac{\pi}{4}\right], \quad (49)$$

$$\phi_{\mathrm{II}}(x)$$

$\phi_{\mathrm{II}}(x)$ joins smoothly to a linear combination of functions of the form of Eq. (24) and $\phi_{\mathrm{I}}(x)$ joins smoothly to a function of the form of Eq. (30) with $v(x)$ and $p(x)$ assuming their more general form, given by Eqs. (23) and (11), respectively.

The connection formula for a solution that is well-behaved in region III is given by

$$\frac{e^{i\pi x/a}}{[v(x)]^{\frac{1}{2}}}\cos\left[\frac{1}{\hbar}\int_x^{x_2}\left(\frac{\pi\hbar}{a}-p(x)\right)dx-\frac{\pi}{4}\right]\leftarrow \phi_{\mathrm{II}}(x)$$

$$\frac{e^{i\pi x/a}}{2(|v(x)|)^{\frac{1}{2}}}\exp\left(-\frac{1}{\hbar}\int_{x_2}^{x}\left|\frac{\pi\hbar}{a}-p(x)\right|dx\right). \quad (50)$$

$$\phi_{\mathrm{III}}(x)$$

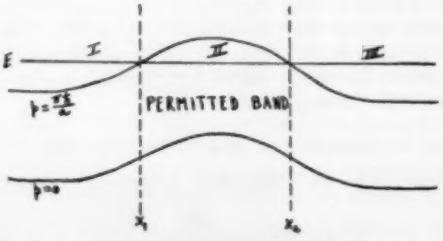


FIG. 4. Impurity state near a $p=\pi\hbar/a$ band edge.

The two expressions for $\phi_{II}(x)$ given in Eqs. (49) and (50) can be made to join smoothly if

$$\int_{x_1}^{x_2}\frac{[\pi\hbar/a-p(x)]}{\hbar}dx=(m+\tfrac{1}{4})\pi, \tag{51}$$

where m is an integer.

IV. STATES IN WHICH A PERMITTED BAND IS CROSSED

Suppose the perturbing potential is of the type indicated in Fig. 5. The connection formula for a solution that is well-behaved in region I is given by

$$\frac{1}{2}\frac{e^{i\pi x/a}}{(|v(x)|)^{\frac{1}{2}}}\exp\left(\frac{1}{\hbar}\int_x^{x_1}\left|\frac{\pi\hbar}{a}-p(x)\right|dx\right)\rightarrow$$

$$\phi_I(x)$$

$$\frac{e^{i\pi x/a}}{[v(x)]^{\frac{1}{2}}}\cos\left[\frac{1}{\hbar}\int_{x_1}^{x}\left(\frac{\pi\hbar}{a}-p(x)\right)dx-\frac{\pi}{4}\right]. \tag{52}$$

$$\phi_{II}(x)$$

The connection formula for a solution that is well-behaved in region III is given by

$$\frac{1}{[v(x)]^{\frac{1}{2}}}\cos\left[\int_x^{x_2}\frac{p(x)}{\hbar}dx-\frac{\pi}{4}\right]\leftarrow$$

$$\phi_{II}(x)$$

$$\frac{1}{2}\frac{1}{(|v(x)|)^{\frac{1}{2}}}\exp\left(-\frac{1}{\hbar}\int_{x_1}^{x}|p(x)|dx\right). \tag{53}$$

$$\phi_{III}(x)$$

The two solutions in region II, namely

$$\phi_{II}{}^{A}(x)=\frac{e^{i\pi x/a}}{[v(x)]^{\frac{1}{2}}}\cos\left[\frac{1}{\hbar}\int_{x_1}^{x}\left(\frac{\pi\hbar}{a}-p(x)\right)dx-\frac{\pi}{4}\right], \tag{54}$$

and

$$\phi_{II}{}^{B}(x)=\frac{1}{[v(x)]^{\frac{1}{2}}}\cos\left[\frac{1}{\hbar}\int_{x}^{x_2}p(x)dx-\frac{\pi}{4}\right], \tag{55}$$

clearly cannot be made to coincide for all values of x. However, since the wave function [Eq. (3)] depends only on the values of ϕ at the lattice points, it need only be required that $\phi_{II}{}^{A}$ and $\phi_{II}{}^{B}$ coincide at these points. At the lattice points, given by $x=na$ where n is an integer, $\phi_{II}{}^{A}$ can be written

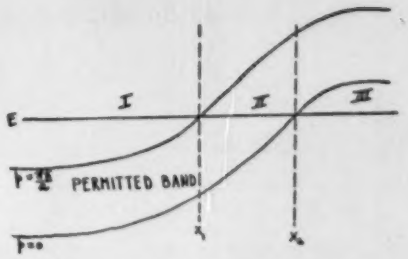


FIG. 5. Energy state in which a permitted band is crossed.

$$\phi_{II}{}^{A}(x)=\cos\left[\frac{1}{\hbar}\int_x^{x_1}p(x)dx-\frac{\pi x_1}{a}-\frac{\pi}{4}\right]$$

$$=\cos\left[\frac{1}{\hbar}\int_x^{x_2}p(x)dx-\frac{\pi}{4}\right.$$

$$\left.-\frac{1}{\hbar}\int_{x_1}^{x_2}p(x)dx-\frac{\pi x_1}{a}\right]. \tag{56}$$

$\phi_{II}{}^{A}(x)$ and $\phi_{II}{}^{B}(x)$ can thus be brought into agreement at the lattice points when

$$\int_{x_1}^{x_2}\frac{p(x)}{\hbar}dx+\frac{\pi x_1}{a}=m\pi, \tag{57}$$

where m is an integer. It should be emphasized that in this formulation the origin of coordinates is at a lattice point. A shift in origin to another lattice point increases x_1 by a multiple of a, thereby increasing the term $\pi x_1/a$ by a multiple of π. Thus the shift in origin does not alter the quantum condition.

For the special case in which a uniform electric field is applied to the crystal, the potential $V(x)$ has the form $V(x)=Fx$. Then $\int_{x_1}^{x_2}[p(x)/\hbar]dx$ is independent of the energy, and it follows easily from Eq. (57) that the permitted energy states are spaced an amount $\Delta E=Fa$ apart. This result can be argued from other considerations.[1]

V. ACKNOWLEDGMENTS

The author would like to express her warmest thanks to Professor H. M. James for many helpful discussions of the problem.

Revised Houston Method for Counting Lattice Frequencies

JENN-LIN HWANG
Department of Physics, Taiwan University, Taipei, Formosa, China
(Received March 8, 1955)

A modified Houston-Nakamura approximation is suggested for counting lattice frequencies. This method is tested on a transversely vibrating two-dimensional lattice and found to give excellent agreement with the exact Bower-Rosenstock solution. Following Houston, the method interpolates between symmetry lines or planes to obtain frequency contours within the Brillouin zone. The difference is that the interpolation does not follow the spherical harmonic approximation of Houston but suggests contours of specified form intermediate between circles and straight lines in two dimensions and between spheres and planes in three dimensions. Nakamura's procedure of using separate approximations in different regions of the Brillouin zone is also followed.

IT was suggested five years ago by Nakamura[1] that the spurious peaks characteristics of Houston's method[2,3] for calculating the frequency distribution of a crystal can be removed by considering the topological behavior of the contours of constant frequency in the reciprocal lattice. In spite of this valuable suggestion, no actual calculation for a three-dimensional lattice has been carried out so far. The reasons may be: (1) Before the calculation one should designate several regions in the half-periodicity cube to be treated in different ways, and the critical points of the frequency usually play an important role in the determination of their boundaries. However no practical method for finding the critical points, except those at corners, has been given. (2) Spherical harmonics, in a two-dimensional case Fourier series, are inadequate for approximating the contours of constant frequency.

The first drawback may be easily removed now. Recently Lax and Lebowitz,[4] and Rosenstock[5] have independently improved the Montroll method of moments[6] by applying the results of Van Hove's analysis.[7] Rosenstock has shown that the location of the critical points may be found by studying the frequencies $[\tilde{\omega}(q)=\omega/\omega_{\max}$ or $\tilde{\omega}^2(q)=W(q)$; hereafter we use the latter] on the edges and surfaces of the half-periodicity cube, where, as was noticed by Houston, the secular equation may be solved analytically. Actually by a simple consideration the behavior inside the cube may be inferred with little ambiguity from such a small number of edges or surfaces, and hence the boundaries of the divided regions may be easily determined. To overcome the second drawback we shall use an interpolating formula [Eq. (1) below] instead of the spherical harmonics. When the secular equation is solved along two directions OA and OB [Fig. 1(a)] making an angle ϑ and the distances from the origin are expressed in terms of the frequency, say $r_1=f(W,0)$ and $r_2=f(W,\vartheta)$, the following two rules hold empirically:

(1) If $r_1=r_2$, the contour of constant frequency is a circular arc, i.e., $r=f(W,\theta)=$const.

(2) If $r_1=r_2\cos\vartheta$, the contour of constant frequency is a straight line perpendicular to OA, i.e., $r=r_1\sec\theta$, and if $r_2=r_1\cos\vartheta$, then it is the one perpendicular to OB,[8] i.e., $r=r_2\sec(\vartheta-\theta)$.

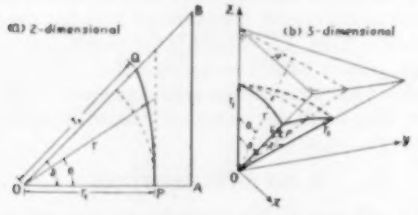


FIG. 1. Approximate contours of constant frequency (a) 2-dimensional, (b) 3-dimensional.

In a divided region of the half-periodicity cube (square) we can generally find either of the contours of such simple shapes at the corner or in the vicinity of the border, and can presume them to be two extreme cases of the contours of constant frequency. The other contours are then approximated by a linear interpolation, or extrapolation if necessary, between these extreme ones. In Fig. 1(a) we suppose $r_1=r_2\cos\vartheta$ at A and B, that is, the contour is a straight line perpendicular to OA, and draw a perpendicular line through P and a circular arc with center O and radius r_1. Then the equation of the approximate contour of constant frequency passing through the two points P and Q, their distances from O being r_1 and r_2 respectively, is given by

$$r=r_1+\frac{r_1\sec\theta-r_1}{r_1\sec\vartheta-r_1}(r_2-r_1),$$

or

$$r=r_1+\frac{\sec\theta-1}{\sec\vartheta-1}(r_2-r_1), \qquad (1)$$

[1] T. Nakamura, Progr. Theoret. Phys. (Japan) **5**, 213 (1950).
[2] W. V. Houston, Revs. Modern Phys. **20**, 161 (1948).
[3] E. Bauer, Phys. Rev. **92**, 58 (1953).
[4] M. Lax and J. L. Lebowitz, Phys. Rev. **95**, 617 (1954); **96**, 594 (1954).
[5] H. B. Rosenstock, Phys. Rev. **95**, 617 (1954); **97**, 290 (1955).
[6] E. W. Montroll, J. Chem. Phys. **10**, 218 (1942); **11**, 481 (1943); E. W. Montroll and D. C. Peaslee, J. Chem. Phys. **12**, 98 (1944); C. W. Garland and G. Jura, J. Chem. Phys. **22**, 1108 (1954).
[7] L. Van Hove, Phys. Rev. **89**, 1189 (1953).
[8] There is an exception to the rule (2) in three-dimensional cases, but such an exception may be easily distinguished.

TABLE I. Comparison of the approximate and exact frequency distributions ($W=\omega^2/\omega_{max}^2$). G_A is our approximate frequency distribution. G_E is the exact distribution of Bower and Rosenstock (reference 9).

	$\sigma=1/4$			$\sigma=1/2$			$\sigma=3/4$	
W	$G_A(W)$	$G_E(W)$	W	$G_A(W)$	$G_E(W)$	W	$G_A(W)$	$G_E(W)$
0.00	0.424	0.424	0.00	0.318	0.318	0.00	0.318	0.318
0.05	0.444	0.444	0.10	0.343	0.344	0.20	0.376	0.381
0.10	0.466	0.467	0.20	0.373	0.375	0.40	0.468	0.483
0.20	0.518	0.519	0.30	0.410	0.415	0.60	0.650	0.688
0.30	0.585	0.587	0.40	0.457	0.465	$0.80_{(-)}$	1.52	1.59
0.40	0.675	0.680	0.50	0.519	0.531	$0.80_{(+)}$	3.11	3.18
0.50	0.805	0.818	0.60	0.606	0.625	$\left[\begin{matrix}0.80\\(+)-(-)\end{matrix}\right]$	1.59	1.59
0.60	1.01	1.05	0.70	0.739	0.768	0.81	3.50	3.57
0.70	1.50	1.61	0.80	0.973	1.02	0.82	4.21	4.19
0.75	∞	∞	0.85	1.18	1.25	5/6	∞	∞
0.80	1.68	1.75	0.90	1.55	1.65	0.84	5.26	4.82
0.85	1.38	1.49	0.95	2.45	2.63	0.85	3.44	3.70
0.90	1.27	1.36	0.98	4.48	4.81	0.90	2.09	2.19
0.95	1.24	1.30	1.00	∞	∞	0.95	1.69	1.69
1.00	1.27	1.27				1.00	1.42	1.42

regardless of the position of Q. This is the basic formula of our approximation.

Applications have been made to the "transverse" vibrational case of a square lattice,[9] where the frequency distributions have already been derived analytically by Bower and Rosenstock. From the viewpoint of approximate methods there is of course no essential difference between this simple example and the "in-plane" vibrational case of other two-dimensional lattices. The secular equation was found to be

$$\omega^2=2\alpha(2-\cos x-\cos y)+4\gamma(1-\cos x\,\cos y)\quad 0<x,\,y<\pi$$

or

$$W=\begin{cases}\frac{1}{2}[\frac{1}{2}(2-\cos x-\cos y)+\sigma(1-\cos x\,\cos y)] \text{ for } \sigma\leq\frac{1}{2}.\\ \dfrac{1}{1+2\sigma}[\frac{1}{2}(2-\cos x-\cos y)+\sigma(1-\cos x\,\cos y)]\\ \qquad\qquad \text{for}\quad \sigma\geq\frac{1}{2},\ \sigma=\gamma/\alpha.\end{cases}\tag{2}$$

Case (i): $\sigma=\frac{1}{2}$.—Solving Eq. (2) along the three sides of the $\frac{1}{8}$-half-periodicity "square," we have $r=\cos^{-1}\times(1-2W)$ for $y=0$, $r=\sqrt{2}\cos^{-1}[2(1-W)^{\frac{1}{2}}-1]$ for $x=y$, and $W=1$ for $x=\pi$. Since the contour $W=1$ is a straight line $x=\pi$, the whole half-periodicity square is a region. Taking $y=0$ as r_1, $x=y$ as r_2, $\vartheta=\pi/4$, and introducing Eq. (1), we obtain the frequency distribution

$$G(W)=\frac{1}{\pi^2}\frac{d}{dW}\int_0^{\pi/4}r^2(W,\vartheta)d\theta\tag{3a}$$

$$=0.15916r_1r_1'+0.04695(r_1\eta_{21}'+r_1'\eta_{21})+0.02675\eta_{21}\eta_{21}',\tag{3b}$$

where $\eta_{21}=r_2-r_1$.

Case (ii): $\sigma=1/4$.—The contour $W=7/9$ is a straight line perpendicular to $x=y$. We divide the square into two regions by the contour $W=3/4$ which passes through the corner $(\pi,0)$.[10] A direct application of Eq. (3b) to each region yields the distribution function shown in Fig. 2(b), in which $r_1=\sqrt{2}\cos^{-1}[3(1-8W/9)^{\frac{1}{2}}-2]$, $r_2=\cos^{-1}(1-8W/3)$ for $0\leq W\leq 3/4$, and $r_1=\sqrt{2}\times\{\pi-\cos^{-1}[3(1-8W/9)^{\frac{1}{2}}-2]\}$, $r_2=\pi-\cos^{-1}(1-8W/3)$ for $3/4<W\leq 1$.

Case (iii): $\sigma=3/4$.—To improve the poor precision, obtained by Nakamura, in the rectangular part bounded by the contour $W=5/6$, $y=0$, and $x=\pi$, further knowledge along the diagonal AP [Fig. 2(c)] should be acquired. However, the secular equation along this line is not so simple. We have fortunately found a good approximation. The secular equation along $y=0$, $x=\pi$,

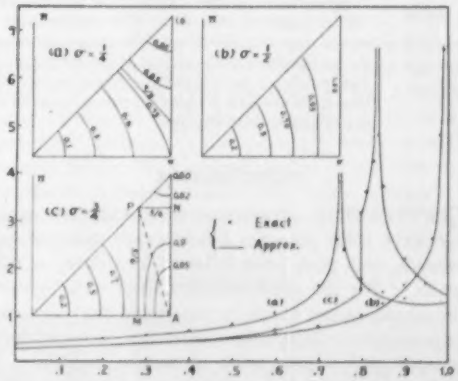


FIG. 2. Contours of constant frequency and frequency distributions for the transversal vibration of a two-dimensional crystal; $W=\omega^2/\omega_{max}^2$. (a) $\sigma=1/4$, (b) $\sigma=1/2$, (c) $\sigma=3/4$; $\sigma=\gamma/\alpha$.

[9] W. A. Bower and H. B. Rosenstock, J. Chem. Phys. **18**, 1056 (1950).

[10] According to Eq. (1) a trivial discrepancy would occur between the different expressions for the border $W=3/4$ in both regions. If one does not prefer this, he may divide the square into three sections $0\leq W\leq 3/4$, $3/4\leq W\leq 7/9$, and $7/9\leq W\leq 1$. In the second region Eq. (1) also holds after a slight modification.

and $y=\pi-x$ may be reduced, except for a scaling factor, to an equation (with A as origin):

$$W=\tfrac{1}{2}(1+W_0)+\tfrac{1}{2}(1-W_0)\cos r,$$

W_0 being the value of W on the diagonal $y=x$. This serves a good approximation for the other direction. Along AP we have $W_0=5/6$ and $r=(l/\pi)\cos^{-1}\times(12W-11)$, where l is the length of AP. The rectangular part (ratio of sides $AN/AM=\tau$) is then divided into two regions by AP. If the rectangle is reduced to $1/\tau$ along its longer side, we have a square and can make an immediate application of Eq. (3b).[11] Thus the distribution function due to the rectangular part may be found from

$$G(W)=\tau\{a(\bar{r}_1\bar{r}_1'+r_2r_2')+b[(\bar{r}_1\bar{\eta}_{01}'+\bar{r}_1'\bar{\eta}_{01})+(r_2\eta_{02}'+r_2'\eta_{02})]+c(\bar{\eta}_{01}\bar{\eta}_{01}'+\eta_{02}\eta_{02}')\},$$

where a, b, c are the numerical coefficients appearing in Eq. (3), and $\bar{r}_1=r_1/\tau$, $\tau_0=\sqrt{2}AM/l$, $\bar{\eta}_{01}=(r_0/\tau_0)-(r_1/\tau)$, $\eta_{02}=(r_0/\tau_0)-r_2$.

The approximate frequency distributions $G_A(W)$ and the exact ones $G_E(W)$ of Bower and Rosenstock are given in Table I and Fig. 2.

The extension of Eq. (1) to a three-dimensional case is straight forward. For example, as shown in Fig. 1(b), the surface of constant frequency below the first critical point (low-frequency part) is approximated by

$$r(W,\vartheta,\varphi)=r_1+\frac{\sec\vartheta-1}{\sec\vartheta_1-1}(\rho-r_1),\quad \rho=r_2+\frac{\sec\psi-1}{\sec\chi-1}(r_3-r_2),$$

where $\chi=\tan^{-1}(\sqrt{2}/2)$, and ϑ and ψ are connected with φ by $\tan\vartheta=\sec\varphi$ and $\tan\psi=(1/\sqrt{2})\tan\varphi$. The result of applying our method to three-dimensional cases will appear later.

[11] The following scheme is somewhat better than that illustrated in the text. By a simple test we find that the ratio $r_1/r_2=\epsilon$ in the neighborhood of A is equal to $\sqrt{5}$. This means that the contour $W\cong 1$ behaves not like a circle but rather like an ellipse. Hence we reduce the rectangle to $1/\epsilon$ along the longer side before applying Eqs. (1) and (3a). Of course, the numerical coefficients are no longer equal to those of Eq. (3b), and they should be evaluated with regards to the different values of ϑ.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Electroluminescence Excited by Short Field Pulses*

FRANK MATOSSI, *United States Naval Ordnance Laboratory, White Oak, Maryland*

AND

SOL NUDELMAN, *United States Naval Ordnance Laboratory, White Oak, Maryland, and University of Maryland, College Park, Maryland*

(Received February 2, 1955; revised manuscript received May 24, 1955)

The decay of the light emitted from several electroluminescent phosphors excited by periodically repeated voltage pulses of from 40 to 2500 microseconds duration has been determined. The slopes of the decay curves after each individual pulse excitation become steeper as the pulse duration decreases. The heights of the luminescence peaks excited at the leading and trailing edges of the pulses also depend on the pulse time. The results are interpreted as confirming previous conclusions about the influence of polarization charges. Numerical evaluation yields the correct order of magnitude for the "critical time" needed for the accumulation of polarization charges.

INTRODUCTION

EXCITATION of electroluminescence by square wave fields produces luminescence peaks at field reversals, each peak being followed by a decay of light output.[1,2] For the green emission band of a Sylvania electroluminescent panel, it was found that this electroluminescence decay is faster than the "natural" decay obtained after steady weak ultraviolet excitation. In the blue band of this panel, however, it is slower than the corresponding natural decay. These and other results obtained with nonsinusoidal fields were interpreted as being caused by the action of polarization charges that pile up near the surface and reduce the effective field.[1]

In order to further test this assumption, phosphors were excited by voltage pulses whose duration could be made sufficiently short so that polarization charges would not accumulate. Furthermore, an attempt will be made to apply the experimental results to recent theories of polarization effects in semiconductors.[3]

METHOD

Repeated voltage pulses were obtained from two pulse generators and then amplified. The first generator

* Based on a dissertation submitted by one of the authors (S.N.) to the University of Maryland in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[1] S. Nudelman and F. Matossi, J. Electrochem. Soc. **101**, 546 (1954).

[2] Zalm, Diemer, and Klasens, Philips Research Repts. **9**, 81 (1954).

[3] J. R. McDonald, Phys. Rev. **92**, 4 (1953); R. J. Friauf, J. Chem. Phys. **22**, 1329 (1954).

produced wide pulses whose duration could be varied down to about 10% of the period (e.g., 500 μsec at 200 cps), with a rise time of about 150 μsec. The second generator provided pulses 40 to 75 μsec wide. Pulses appreciably shorter than 40 μsec could not be used since the finite rise time interfered. The pulse repetition rate was usually maintained at 200 cps. The potential differences applied were between 300 and 1000 volts.

The light output was observed with a photomultiplier in connection with a direct-coupled oscilloscope. Appropriate filters[1] separated the green and blue emission bands.

The phosphors were used as powdered material embedded in a plastic (parlodium) and subjected to the field in cells of the type described previously.[4] Different kinds of electroluminescent phosphors were investigated. The results obtained are similar for all of them. Details will, therefore, be reported only for the green Sylvania phosphors, which were studied most extensively in the form of Sylvania Panelecent Lamps and as phosphors made according to the method given by Homer *et al.*[5] The results were essentially independent of the voltage and also independent of whether the pulses were applied unilaterally or by reversing the direction of the field.

RESULTS

Figure 1 shows a selection of oscillograms for the Sylvania phosphors. The difference between the curves for the blue and the green band is quite obvious. The results obtained with the wide pulses have already been reported.[1] The oscillograms for the narrow pulses readily show that the decay has become faster.

A so-called "blue" Sylvania panel with a considerably slower blue natural decay (50% of initial value in about 10 msec) than that of the green Sylvania Lamp (50% in about ½ msec) showed, in the blue region, a pattern resembling the "green" patterns of Fig. 1.

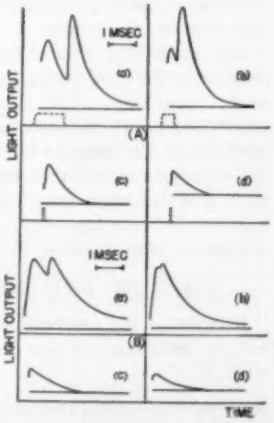


FIG. 1. Oscillograms of luminescence of a Green Sylvania Lamp excited by electric field pulses. (A) Blue luminescence: (*a*) 1250, (*b*) 500, (*c*) 75, (*d*) 40 μsec pulse width. (B) Green luminescence: (*a*) 1000, (*b*) 500, (*c*) 75, (*d*) 40 μsec pulse width. – – – – Shape of voltage pulse. The oscillograms show the luminescence intensity *versus* time in one period of the field. The solid horizontal line represents zero light output.

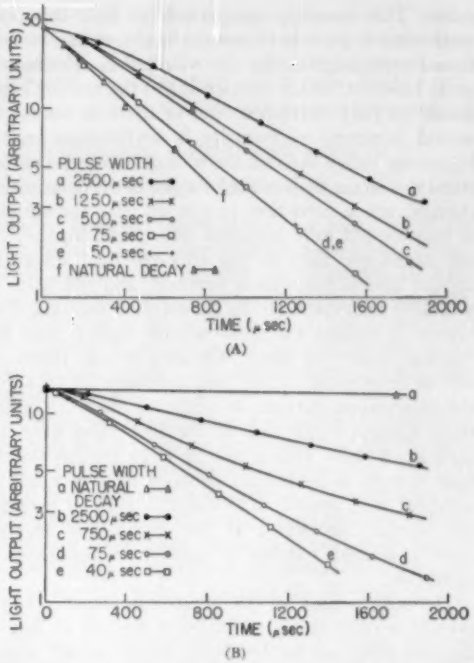


FIG. 2. Decay curves for green Sylvania phosphor. (A) Blue luminescence: (a)–(e) electroluminescence decays for different pulse widths, (f) natural decay. (B) Green luminescence: (a) natural decay, (b)–(e) electroluminescence decays. The decay curves show the decay of intensity after the second peak of Fig. 1. The origin of the time scale is the time of the maximum of the luminescence peak.

Quantitative details are shown in Fig. 2. With decreasing pulse width, the blue decay becomes more and more rapid, and approaches the natural one shown in curve *f* of Fig. 2(A). The green decay [Fig. 2(B)] also becomes steeper with shorter excitation times and, therefore, deviates in the direction away from the natural decay. With respect to the decay curves, all phosphors behaved qualitatively alike although the natural decays varied appreciably, particularly in the green band.

DISCUSSION

The oscillograms obtained with the wide pulses can be interpreted by considering the decay times of the natural decays and the efficiency of excitation.[1] That indeed the natural decay times are controlling factors could be seen from the behavior of the blue Sylvania panel, which yielded patterns like those of Fig. 1(B) because of similar decay times.

The inspection of the oscillograms of Fig. 1 shows that the first peak, excited by the leading edge of the pulse, has practically disappeared at the very narrow

[4] F. Matossi and S. Nudelman, Phys. Rev. **89**, 660 (1953).
[5] Homer, Rulon, and Butler, J. Electrochem. Soc. **100**, 566 (1953). Details about the phosphors used are described in S. Nudelman's thesis, University of Maryland, 1955 (unpublished).

pulses. This is readily understandable since this peak needs time to grow to its normal height, as can be seen from the oscillograms for the wide pulses. The second peak, however, which corresponds to the trailing edge, should be fully developed. But in the blue band, the second increment of intensity is nevertheless smaller for narrow pulses than for the wide ones. We can understand this on the basis of the behavior of the polarization charges, which need time to pile up near the surface. If they do not have sufficient time to do so, they do not contribute fully to the effective field when the applied field is removed or reversed, and this leads to a smaller increment at the second field reversal. This begins to happen for pulse widths smaller than the "critical time"[1] of about 700 μsec, which, therefore, will be interpreted as the time necessary for piling up the polarization charges. In addition, the finite, albeit small natural decay time of the blue band may, for such small pulse widths, not be large enough for the blue centers to be refilled between field reversals in sufficient numbers. This would also tend to decrease the intensity increment at the second field reversal since a smaller number of centers is available for excitation.

The second peak of the green luminescence is expected[1] to reach the same height as the first as long as the first peak can be developed to its normal height. This is true for the frequencies around 100 cps. But there are deviations at other frequencies. At higher frequencies, the second peak is a little higher, as in Fig. 1(B*b*); at lower frequencies, it is a little smaller. This detail is not yet understood.

The green decay time is controlled by the emptying of traps filled during excitation (phosphorescence decay). It can readily be assumed that field excitation will not fill as many traps as ultraviolet excitation, and that the number of traps filled will also decrease with decreasing pulse width. Therefore, the electroluminescence decay will be faster than the natural decay and will become still faster with decreasing pulse width.

The decay of the blue luminescence [Fig. 2(A)] is assumed to be intimately connected with the decay of the internal field since the blue emission band, presumedly, does not involve transitions to or from the conduction band or traps. Thus the exciting field and not the filling of traps is the main controlling factor for the blue luminescence of phosphors like the green Sylvania phosphors. In the next section, this relation between electroluminescence decay and field decay shall be discussed in a semiquantitative manner.

DECAY OF THE INTERNAL FIELD

The internal field can be determined from Poisson's equation

$$\partial E/\partial x=\beta(p-n), \quad \beta=4\pi e/\epsilon, \tag{1}$$

if the charge distribution is known. (Here ϵ is the dielectric constant and x is the space coordinate in the direction of the field.) This distribution may be obtained from the equations[3]

$$\begin{aligned}\partial p/\partial t&=-\mu E(\partial p/\partial x)-\mu\beta p(p-n)\\&\quad+D\partial^2 p/\partial x^2+\kappa(c_0^2-pn),\\ \partial n/\partial t&=\mu E(\partial n/\partial x)+\mu\beta n(p-n)\\&\quad+D\partial^2 n/\partial x^2+\kappa(c_0^2-pn),\end{aligned} \tag{2}$$

where p and n are the numbers of holes (or empty centers) and free electrons, respectively; μ is the mobility, D is the diffusion coefficient, both of which are assumed to be equal for holes and electrons, which of course is an approximation only (but other assumptions lead to similar results); E is the field; κ is a transition probability, and c_0 is the number of unexcited centers; the relation $\mu/D=e/kT$ will be assumed to be valid.

These equations adequately describe the influence of polarization charges on capacitance effects in semiconductors,[3] and they could also be utilized, with the addition of a term accounting for periodic excitation by a sinusoidal field, for the interpretation of polarization effects on electroluminescence brightness waves.[4] For the purpose of the present paper, we furthermore assume

$$\partial(p+n)/\partial x=0, \tag{3a}$$

which is approximately true in the case of sinusoidal fields, as can be inferred from the solutions given by Friauf and McDonald.[3] A more restrictive assumption is

$$\partial(p+n)\partial t=0, \tag{3b}$$

so that $p+n=2n_0=$const. This is not rigorously true, but we may safely assume that a loss of electrons and holes because of recombination may be neglected near the surface and for times short with respect to the duration of the green luminescence, which is controlled by such recombinations, and which is, as is known from the observation of the green natural decay, much longer than the blue decay time.

For $x=0$ and $x=L$ ($L=$thickness of phosphor particle) at $t=0$, we assume $p=2n_0$, $n=0$ and $n=2n_0$, $p=0$, respectively. The situation described by the above conditions corresponds to polarization charges piled up at the surfaces by the external field. These space charges extend into the interior of the particle with a distribution determined by Eq. (2). If the external field is removed, the charge distribution changes in time according to

$$\partial(p-n)/\partial t=D\partial^2(p-n)/\partial x^2-2\mu\beta n_0(p-n). \tag{4}$$

Because of the assumptions (3), the considerations will be restricted to one surface ($x=0$) only. The result is

$$E=2\beta n_0\exp(-\rho Dt)\{1+\cosh[(C-\rho)^{\frac{1}{2}}L]\}/\times(C-\rho)^{\frac{1}{2}}\sinh[(C-\rho)^{\frac{1}{2}}L], \tag{5}$$

where $C=2\mu\beta n_0/D$, $\rho=$const.

[4] F. Matossi, Phys. Rev. **98**, 434 (1955).

The influence of the amount of polarization charges is expressed by the dependence on n_0. The decay of the intensity can be related to the decay of the field by taking into account the field dependence of the intensity, which can be described, with sufficient accuracy, by the empirically well established relation[1,7]

$$I = \text{const} E^2 \exp(-b/E), \tag{6}$$

from which

$$\sigma \equiv d \ln I/dt = -\rho D(2+b/E). \tag{7}$$

In order to apply Eq. (7) to the observations, some numerical estimates are necessary. Since the observed slopes are constant over a substantial time range, it must be assumed that E is nearly constant over the same time range, or that ρDt is sufficiently small. This time range is in the order of 1 msec, which leads to the condition $\rho D \ll 10^3$ sec^{-1}. On the other hand, ρD should not be much smaller than 100 sec^{-1} since otherwise the decay should be expected to extend to times larger than 1/100 sec. Further reasonable numerical assumptions ($n_0 \cong 10^{17}$ cm^{-3}, e.g.) give the order of magnitude of C as $C \cong 10^{13}$ cm^{-2}. Equation (5) is a useful expression for E only if $C > \rho$. This condition can be satisfied, and it may even be assumed $C \gg \rho$, if D is not too small, say $D \gg 10^{-11}$ cm^2/sec. Experimental values for D, are, however, not known.

Applying these numerical estimates, we obtain for the ratio of the slopes for two different amounts of polarization charges,

$$\sigma_1/\sigma_2 = E_2(2E_1+b)/E_1(2E_2+b), \tag{8}$$

where $E \propto n_0^{\frac{1}{2}}$. For very high fields ($E \gg b$), $\sigma_1/\sigma_2 = 1$; for very small fields,

$$\sigma_1/\sigma_2 = E_2/E_1 = (n_{02}/n_{01})^{\frac{1}{4}}. \tag{8a}$$

This ratio is larger than unity if $n_{02} > n_{01}$. Therefore, the slope increases with decreasing amount of polarization charges. It can be shown that the qualitative aspect of this result is not changed by less stringent approximations.

It is possible to utilize Eqs. (8) for a numerical evaluation, if it is assumed that the increase of polarization charges near the surface follows an exponential law like

$$n_0 = \bar{n}_0[1-\exp(-at)], \tag{9}$$

which is equivalent, in our approximation, to

$$E = \bar{E}[1-\exp(-at)]^{\frac{1}{2}}.$$

From the observed slopes of the blue band, which are given in Table I, the time constant a can be computed.

TABLE I. Slopes of blue decay.

Pulse width (μsec)	50	75	500	1250	2500	natural
$-\sigma(10^3$ sec$^{-1})$	0.836	0.836	0.735	0.634	0.565	1.02

From $\sigma_{500}/\sigma_{2500}$ and Eq. (8a), we obtain $a = 1.87 \times 10^3$ sec^{-1}. Furthermore, σ_{1250} can be calculated as $\sigma_{1250} = 0.60$. The calculated slopes for the very narrow pulses come out much too high, indicating that the natural decay interferes and prevents the observation of the influence of the field decay.

Since, in the experiments, the applied voltage was 320 volts and b could be determined as $b \cong 500$ volts (from the field dependence of intensity), Eq. (8) should have been used instead of Eq. (8a). But then the initial value of E near the surface must be known, which is certainly larger than the average field given by the applied voltage. Under these circumstances, a becomes smaller but the order of magnitude is retained. For $\bar{E} = 2\langle E \rangle_{\text{av}}$, we would obtain $a = 1.3 \times 10^3$ sec^{-1}.

The values obtained for a have the correct order of magnitude if we identify $1/a$ with the "critical time" of 700 μsec mentioned in the previous section. The value of a varies somewhat with the selection of the slopes used for the calculation. In part, this variation is the result, of course, of the approximations in the formulas and to the limitations of the experiment, and in part to the disturbing influence of the natural decay.

The qualitative agreement with the observations (steeper slopes with decreasing amount of polarization charges) and the correct order of magnitude of the numerical evaluation may be considered as sufficient indication that the consideration of moving and diffusing polarization charges according to Eqs. (1) and (2) together with Eq. (9) gives an adequate description of the behavior of the internal field.

ACKNOWLEDGMENTS

The authors would like to thank Dr. E. Meschter, R. I. du Pont de Nemours, Parlin, New Jersey, and Dr. J. F. Waymouth, Sylvania Electric Products Company, Salem, Massachusetts, as well as the respective companies for supplying some of the material used and for information about the phosphors.

[7] G. Destriau, Phil. Mag. **38**, 700 (1947). Other empirical relations, except pure power laws, would not alter the general conclusions.

Thermal Conductivity of Germanium at Ambient Temperatures*

KATHRYN A. MCCARTHY AND STANLEY S. BALLARD†
Tufts College, Medford, Massachusetts
(Received March 14, 1955)

The thermal conductivity of high-purity germanium has been measured; it varies approximately linearly from 0.61 watt (cm C°)$^{-1}$ at 5°C to 0.50 watt (cm C°)$^{-1}$ at 95°C.

THE thermal conductivity of high-purity germanium has been measured through the temperature range 5 to 95°C, employing a technique[1] developed especially for use with small samples of poorly conducting materials such as optical crystals. The method is comparative; the thermal conductivity of the "unknown" is measured against a "standard" material of well-established thermal conductivity. In the present measurements, the comparison materials were cast zinc[2] and nickel,[3] both of high purity, obtained from the New Jersey Zinc Company and the International Nickel Company, respectively.

The germanium samples were furnished by the Westinghouse Research Laboratories of East Pittsburgh, Pennsylvania. It was stated[4] that they were cut from an oriented ingot of *n*-type germanium of resistivity about 40 ohm-cm. Two samples were tested, one of which was a 1-cm cube and the other a parallelepiped with dimensions 1 cm×1 cm×0.5 cm, oriented for heat flow perpendicular to the short side. In both cases the orientation of the samples was such that the heat flow was parallel to a 100 crystal direction.

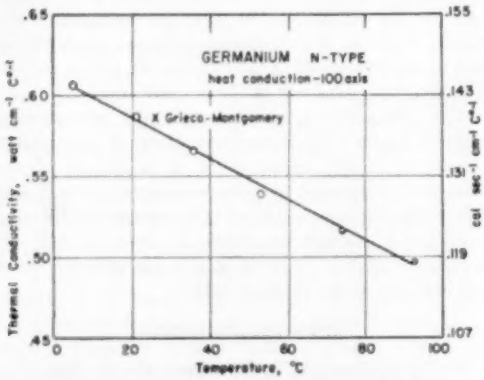


FIG. 1. New thermal conductivity data (plotted as circles) for high-purity *n*-type germanium. The value previously reported by Grieco and Montgomery is also plotted, as ✖. Abscissa units are watt (cm C°)$^{-1}$ on the left, and cal (sec cm C°)$^{-1}$ on the right.

In Fig. 1, the data are plotted as circles, and a straight line has been drawn through them somewhat arbitrarily. It is noted that in this temperature range the thermal conductivity decreases with increasing temperature—presumably this trend might be reversed at temperatures approaching the melting point of germanium (960°C) because of the increasing role then played by electron conduction.

The values of thermal conductivity involved in these measurements are at the very top of the range for which the equipment was designed, but it is still felt that the reproducibility of the individual data points is between 5 and 10 percent. The temperature difference across the sample during measurement was about 4°C. Since this is a small value and the variation of thermal conductivity with temperature is not great, it appears justifiable to take the average temperature between the two faces of the crystal as the temperature for which data are quoted.

A value of the thermal conductivity of germanium, resistivity 10 ohm-cm, was reported in 1952 by Grieco and Montgomery[5]; their one data point is plotted in Fig. 1. They stated additionally that the thermal conductivity was found to be roughly 20 percent less at 100°C—this would give a point lying somewhat below the extension of our straight-line curve. Their determinations were made by a rather similar comparative method, and employed germanium crystals grown at the Bell Telephone Laboratories and apparently of purity not too different from that of our samples. Thus the two sets of data appear to be in substantial agreement.

ACKNOWLEDGMENTS

Appreciation is expressed to our colleague David W. MacLeod for his valuable assistance in the design, fabrication, and maintenance of the equipment, and to the three laboratories that furnished the samples used. The apparatus was constructed with the assistance of a grant received by one of us (SSB) from the Penrose Fund of the American Philosophical Society.

* Presented at the Berkeley meeting of the American Physical Society, on December 29, 1954 [Phys. Rev. **98**, 271 (1955)].
† Present address: Scripps Institution of Oceanography, University of California, La Jolla, California.
[1] Ballard, McCarthy, and Davis, Rev. Sci. Instr. **21**, 905 (1950).
[2] National Bureau of Standards, Circular No. 395, 1931, pp. 69–71.
[3] E. N. Skinner (private communication).
[4] R. E. Davis (private communication).
[5] A. Grieco and H. C. Montgomery, Phys. Rev. **86**, 570 (1952).

Infrared Quenching and a Unified Description of Photoconductivity Phenomena in Cadmium Sulfide and Selenide

RICHARD H. BUBE
Radio Corporation of America, RCA Laboratories Division, Princeton, New Jersey
(Received April 27, 1955)

Infrared quenching, superlinearity, temperature dependence, and slow *S*-shape growth of photocurrent in cadmium sulfide and cadmium selenide crystals may all be associated with the filling and emptying of a specific set of defect levels, as indicated by the location of the calculated electron Fermi level. Infrared quenching of photoconductivity occurs only when the Fermi level lies less than 0.8 ev from the conduction band in CdS, or less than 0.6 ev in CdSe; superlinear photoconductivity occurs only when the Fermi level varies between 0.6 and 0.8 ev from the conduction band in CdS, or between 0.3 and 0.6 ev in CdSe; the photosensitivity decreases rapidly as the Fermi level drops further than 0.6 ev from the conduction band in CdS with increasing temperature, or further than 0.3 ev in CdSe; a slow, *S*-shaped growth of photocurrent is observed when the Fermi level rises between 0.8 and 0.6 ev from the conduction band in CdS, or between 0.6 and 0.3 ev in CdSe.

Infrared quenching maxima for photon energies of 1.65, 1.35, and 0.89 ev are found for CdS, and 1.20, 1.05, and 0.79 ev for CdSe. The magnitudes of the two lower-energy quenching maxima have a constant ratio to one another (as determined from measurements on CdS), as they vary in different crystals, and are most prominent in photosensitive crystals.

INTRODUCTION

ROOM temperature measurements of (1) infrared quenching of photoconductivity and (2) photocurrent as a function of light intensity, for cadmium sulfide and cadmium selenide crystals, lead to an apparent paradox. The concepts of recombination processes advanced by Rose[1] indicate that both infrared quenching and superlinear photoconductivity (i.e., photocurrent increasing with light intensity according to a power greater than unity) should be associated with the filling and emptying of the same set of levels. It is therefore expected either that both phenomena would occur in a given crystal, or that neither phenomenon would occur. Room temperature measurements, however, show that generally CdS crystals show infrared quenching, but not superlinear photoconductivity[2]; crystals of CdSe show superlinear photoconductivity, but not infrared quenching.

It is one of the purposes of this paper to demonstrate that this apparent paradox can be resolved in essential agreement with the general concepts of recombination processes which have been developed. Further data are also provided to aid in correlating these phenomena with the properties of crystal defects.

In brief, it is shown that the superlinear photoconductivity associated with the centers which give rise to infrared quenching in CdS at room temperature, can be detected at elevated temperatures; the infrared quenching associated with the centers which give rise to superlinear photoconductivity in CdSe at room temperature, can be detected at lower temperatures.

Another purpose of this paper is to show that a close correlation exists between infrared quenching, superlinearity, temperature dependence, slow *S*-shape growth of photocurrent in CdS and CdSe, and the calculated location of the quasi- or steady-state electron Fermi level.[3,4] This Fermi level, E_f, is calculated from the conductivity and the temperature according to the equation

$$E_f = kT \ln(N_e e\mu/\sigma), \tag{1}$$

where T is the absolute temperature, N_e is the concentration of states in the lowest kT-wide part of the conduction band, e is the electronic charge, μ is the mobility (assumed to be 100 cm²/volt sec),[5] and σ is the conductivity. $\sigma/e\mu$ is simply the concentration of conduction electrons. A similar hole Fermi level can be defined in terms of the concentration of free holes and the concentration of states in the highest kT-wide part of the valence band.

In describing electronic phenomena, it is convenient to consider these Fermi levels as indicating the approximate boundary between shallow trapping centers, lying above the electron Fermi level or below the hole Fermi level, and recombination centers, lying between the two Fermi levels. If a level does not lie between the two Fermi levels, it is assumed that it is in thermal equilibrium with the nearest allowed band, and does not play an appreciable role in recombination processes. The actual boundary between trapping and recombination centers, sometimes called the demarcation level,[6]

[1] A. Rose, Phys. Rev. **97**, 322 (1955).

[2] This paper discusses results obtained principally with "pure" CdS and CdSe crystals. In evaporated or sintered layers of CdS [see S. M. Thomsen and R. H. Bube, Rev. Sci. Instr., **26**, 664 (1955)] with high proportions of Cu [100–1000 parts per million (ppm)], it is possible to observe a superlinear photoconductivity at room temperature, associated with other levels introduced by the Cu. Photoconductivity phenomena in such CdS:Cu layers is very similar in many respects to that found in CdSe "pure" crystals.

[3] A. Rose, RCA Rev. **12**, 362 (1951).

[4] R. H. Bube, J. Chem. Phys. **23**, 18 (1955).

[5] R. W. Smith, RCA Rev. **12**, 350 (1951).

[6] An electron at the electron demarcation level has equal probability of being thermally excited to the conduction band and recombining with a hole in the valence band. A hole at the hole demarcation level has equal probability of being thermally excited

differs from the Fermi level by a factor of kT times the natural logarithm of the ratio of filled recombination centers to empty recombination centers.[1,7] The difference between Fermi level and demarcation level is usually small; in this paper we shall speak of the Fermi level as if it were also the demarcation level. Unless specified otherwise, it is the electron Fermi level which is referred to in this paper.

Two main pieces of experimental evidence have already been advanced to support the validity of the steady-state Fermi level concept as outlined above: (1) the conductivity at which superlinearity sets in or ceases, with increasing light intensity in CdS or CdSe, varies with the operating temperature in such a way as to give constant values of E_f for the setting in or ceasing of superlinearity,[8] and (2) thermally stimulated current curves obtained for different heating rates have varying magnitudes and temperature locations so as to give constant values of E_f for the current maxima.[4] This paper presents a number of other experimental evidences of a similar nature.

The occurrence of infrared quenching of photoconductivity in CdS and ZnS has been reported by several investigators.[9–11] The most detailed investigation of the nature of the phenomenon to date has been reported by Taft and Hebb.[12] They found two quenching bands with maxima at 0.9 and 1.5 ev, which they associated with excitation of trapped holes in two different types of hole trapes. They also reported that it was possible to attribute the dynamic phenomena involved in an observation of rise and decay of quenching to different time constants in the excitation and quenching processes caused by the same wavelength.

Superlinear photoconductivity in CdSe[5,13] has been investigated as a function of temperature.[8] It was found that superlinearity occurred only when the Fermi level varied between 0.3 and 0.6 ev from the conduction band. Similar measurements for CdS[8] showed that superlinearity occurred at elevated temperatures, when the Fermi level varied between 0.6 and 0.8 ev from the conduction band.

According to the hypothesis advanced by Rose,[1] the explanation of superlinearity and infrared quenching of photoconductivity requires the assumption that there are two classes of recombination centers distributed through the forbidden gap. One possible example is that (1) Class II centers have a capture cross section for holes equal to or greater than that of Class I centers; (2) Class II centers have a capture cross section for electrons less than that of Class I centers, once a hole has been captured; and (3) Class II centers lie below the middle of the forbidden gap and are filled with electrons in the dark. Crystal defects that have the capture cross-section properties stipulated can easily be visualized, purely on electrostatic grounds:

Class I centers:

$$\begin{aligned}[V_a^-]^+ + h^+ &\rightarrow [V_a]^{+\,+},\\ [V_a^{-\,-}]^0 + h^+ &\rightarrow [V_a^-]^+.\end{aligned} \tag{2}$$

Class II centers:

$$\begin{aligned}[V_c]^{-\,-} + h^+ &\rightarrow [V_c^+]^-,\\ [V_c^+]^- + h^+ &\rightarrow [V_c^{+\,+}]^0.\end{aligned} \tag{3}$$

V_c represents a cation vacancy and V_a an anion vacancy. The sign inside the bracket represents the number of trapped electrons or holes; the sign outside the bracket represents the effective charge of the defect with respect to the rest of the crystal.

It is assumed further that Class II centers are distributed over a certain definite energy range in the forbidden gap, located so that at high temperatures and/or low excitation intensities (i.e., for Fermi-levels near the center of the gap), Class II centers lie below the hole Fermi level and only Class I centers act as recombination centers. Then an increase in light intensity over the proper range will lower the hole Fermi level (and raise the electron Fermi level), enable some of the Class II centers to act as recombination centers, and hence cause a shift in the recombination traffic from Class I to Class II centers, i.e., most of the holes formed by excitation will be located in Class II centers, and most of the electrons initially in Class II centers will be located in Class I centers. The effect is to increase the lifetime of a free electron and therefore to increase the photosensitivity. As long as the increasing light intensity lowers the hole Fermi level to include more Class II centers, a superlinear photoconductivity will be observed. (These statements rest on the assumption that the concentration of free electrons and holes is much smaller than the concentration of recombination centers.)

When such a crystal is in a "sensitized" state, i.e., when all or most of the Class II centers are acting as recombination centers, infrared quenching can occur by

---

to the valence band and recombining with an electron in the conduction band. A demarcation level must be defined separately for each type of recombination center as characterized by the ratio of capture cross section for electrons to that for holes.

[7] There is a useful relationship between the electron Fermi-level and the hole demarcation level. The hole demarcation level will lie the same distance above the valence band as the electron Fermi-level lies below the conduction band, with a correction equal to kT times the natural logarithm of the ratio of capture cross section for holes to that for electrons (exactly true only if the concentration of levels in the lowest kT-wide part of the conduction band is the same as the concentration of levels in the highest kT-wide part of the valence band). The sign of the correction is such that the hole demarcation level lies further from the valence band than the electron Fermi level lies from the conduction band if the cross section for holes is greater than the cross section for electrons.

[8] R. H. Bube, in *Proceedings of the Conference on Photoconductivity*, Atlantic City, 1954 (John Wiley and Sons Inc., New York, 1955).

[9] A. E. Hardy, Trans. Am. Electrochem. Soc. **87**, 355 (1945).

[10] R. Frerichs, Phys. Rev. **72**, 594 (1947).

[11] H. W. Leverenz, *An Introduction to Luminescence of Solids* (John Wiley and Sons, Inc., New York, 1950), pp. 164, 302–304, 396.

[12] E. A. Taft and M. H. Hebb, J. Opt. Soc. Am. **42**, 249 (1952).

[13] H. J. Dirksen and O. W. Memelink, Appl. Sci. Research **4B**, 205 (1954).

raising electrons, either from the filled band or from low-lying levels, into the Class II centers. Thus the shift in recombination traffic is reversed, and the photosensitivity is decreased.

The experimental results reported in this paper are used to investigate the problem involved in the determination of the actual location of the defect levels which are responsible for the increased sensitivity responsible for such phenomena as superlinearity, infrared quenching, etc. In the example just discussed above, it was assumed that these levels lay near the filled band. Much of the experimental evidence could be interpreted to indicate that these levels might lie near the conduction band. A discussion of these two possibilities is presented later in the paper.

EXPERIMENTAL

Measurements were made in the photoconductivity apparatus previously described,[4] with which the temperature of a crystal can be varied from −196° to 200°C in an atmosphere of helium. Single crystals of cadmium sulfide and cadmium selenide,[14] prepared by S. M. Thomsen and C. J. Busanovich, were used for the measurements, electrical contact being made using melted indium electrodes.[15]

The bias excitation (hereafter called the primary radiation) for the measurements of infrared quenching was obtained from a GE 1493 incandescent lamp, passed through a 5350 A Farrand filter for measurements with CdS, and through a 6920 A Farrand filter for measurements with CdSe.[16] The primary light flux on the crystal used in these experiments lay in the range between 10^{12} and 10^{13} photons/sec. The quenching radiation (hereafter called the secondary radiation) was obtained from a 500-mm Bausch and Lomb monochromator, used to give 50 A resolution. The secondary flux on the crystal was approximately 10^{14} photons/sec. The photocurrent was recorded on a Leeds and Northrup *X-Y* recorder.

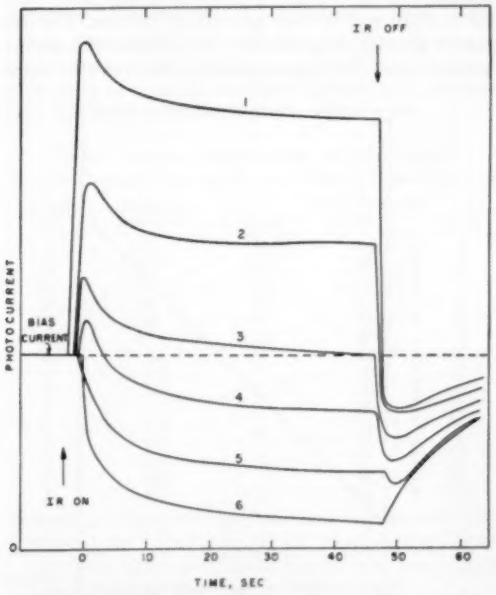


FIG. 1. Dynamic quenching curves for a crystal of CdS for secondary irradiation by wavelengths (1) 6500 A, (2) 6750 A, (3) 7000 A, (4) 7250 A, (5) 7500 A, and (6) 8000 A.

[14] R. H. Bube and S. M. Thomsen, J. Chem. Phys. **23**, 15 (1955).
[15] R. W. Smith and A. Rose, Phys. Rev. **92**, 857(A) (1953).

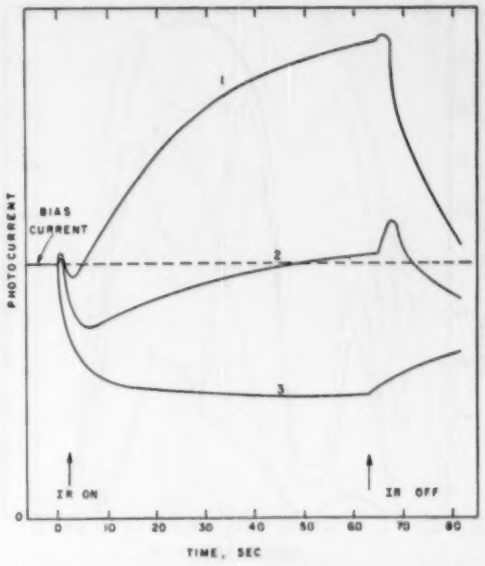


FIG. 2. Dynamic quenching curves for a crystal of CdS for secondary irradiation of wavelengths (1) 7500 A, (2) 7750 A, and (3) 8000 A.

RESULTS

Dynamics of Infrared Quenching

A dynamic method of measurement of infrared quenching was used in this investigation, i.e., the complete curve of photocurrent *vs* time for the onset, equilibrium condition, and end of infrared quenching was recorded for each desired secondary wavelength. Typical results, very similar to those reported by Taft and Hebb,[12] are given in Fig. 1, which is a tracing of actual data obtained for a CdS crystal. The occurrence of a transient maximum current indicates the presence of both excitation and quenching by the secondary radiation. For wavelengths shorter than 6500 A, the current increases asymptotically to an equilibrium value, showing no maximum. As the wavelength of the secondary radiation is increased above 6500 A, the ratio of

[16] It was established that the quenching spectrum is relatively independent of primary radiation wavelength to a first approximation.

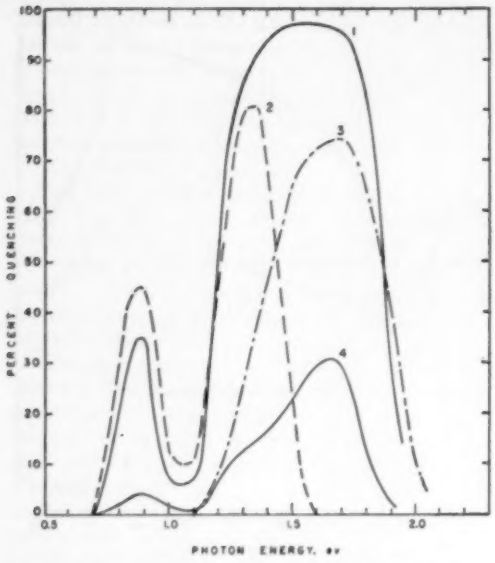


FIG. 3. Infrared quenching spectra at room temperature for four typical crystals of CdS.

excitation to quenching caused by the radiation decreases, until above 7500 A, only quenching remains. Over a broad region of more than a thousand angstroms, however, the secondary radiation causes both excitation and quenching. The form of the curves of Fig. 1 can be explained as was done by Taft and Hebb, by assuming that the excitation caused by the secondary radiation has a higher speed of response than the quenching. When the radiation is removed, the excitation stops very rapidly, but a recovery time is required to remove the quenching effects.

Percent quenching as used in this paper refers to either (1) the ratio of the difference between the equilibrium current and the maximum current, to the maximum current, in cases where the secondary radiation causes some excitation, or (2) the ratio of the difference between the equilibrium current and the bias current, to the bias current, in cases where the secondary radiation does not produce any excitation.

The fact that the form of the dynamic quenching curves can be explained in terms of differing speeds of response for excitation and quenching is dramatically illustrated in Fig. 2. Figure 2 presents a tracing of data obtained for another crystal of CdS, which had a particularly slow excitation speed of response due to trapping effects. Except for the first small increase in current, the curves of Fig. 2 are almost the inverse of the curves of Fig. 1. The curves of Fig. 2 can be explained by assuming that the quenching speed of response of the secondary radiation is higher than the excitation speed of response. The small increase in current found at the turning on of the secondary radiation may be associated with a brief stimulation of electrons from traps.

Infrared Quenching Spectra

It has been previously reported[8] that sensitive CdS crystals show infrared quenching with maximum at about 9000 A, whereas insensitive CdS crystals show quenching with maximum at about 7500 A. These results are explained when it is realized that the main quenching band in CdS consists, not of one band as indicated by Taft and Hebb, but rather of two bands. Figure 3 gives typical infrared quenching spectra for four selected crystals of CdS. Table I summarizes the results obtained with fifteen crystals of CdS, indicating the correlation between the sensitivity of the crystals and the type of infrared quenching spectrum obtained. These measurements were made with constant conductivity produced by the primary radiation. Infrared quenching maxima are found at 0.89 ev and at about 1.35 ev and 1.65 ev.

Some very insensitive crystals of CdS do not show any quenching at all. Insensitive crystal of CdS show predominantly only the 1.65-ev quenching maximum, with half-width of about 0.4 ev. Some relatively insensitive crystals show both the 1.65-ev. and 1.35-ev (half-width of about 0.3 ev) quenching maxima, with about equal magnitudes. The 0.89-ev peak (half-width of about 0.2 ev) occurs when and only when the 1.35-ev peak is found. Sensitive crystals generally show only the 1.35-ev and 0.89-ev quenching maxima. The apparent absence of quenching for photons with energy greater than 1.5 ev in sensitive crystals cannot be taken

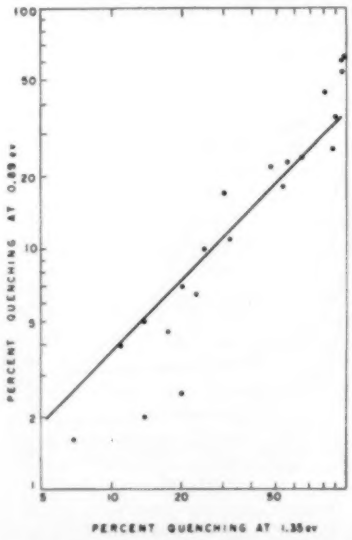


FIG. 4. Percent quenching for photons of 0.89 ev plotted as a function of the percent quenching for photons of 1.35 ev for many different CdS crystals at room temperature.

necessarily as an indication that no quenching occurs in this range. It is possible that the excitation caused by these photons in sensitive crystals overwhelms the quenching effect.[17]

The magnitude of the quenching for the 0.89-ev peak in CdS is plotted as a function of the magnitude of the quenching for the 1.35-ev peak in Fig. 4, for the crystals listed in Table I. Most of the points fall on a straight line with unity slope, indicating a linear relation between the magnitude of quenching, Q, for 0.89-ev and 1.35-ev radiation:

$$Q(0.89\text{ ev}) = 0.37\,Q(1.35\text{ ev}). \quad (4)$$

Infrared quenching spectra for three crystals of pure CdSe at −189°C are given in Fig. 5. Three maxima are found for quenching with photons with energies of 1.20, 1.05, and 0.79 ev. Three major differences between quenching in CdS and CdSe are found: (1) CdSe must be cooled in order to obtain measurements of quenching, (2) all crystals of CdSe tested to date have about the same quenching spectrum, and (3) quenching is much stronger in CdSe than in CdS. Since an incandescent source was used to provide the quenching radiation, some correction of the quenching curve for CdSe is required to account for a decrease in the number of photons per second as a function of energy below 0.9 ev. Between 0.9 and 2.0 ev, the output of the monochromator in photons per second as a function of energy is relatively constant. Because of the lack of secondary radiation intensity for photons with energy below 0.9 ev, the possibility of another quenching maxima in this range for CdSe remains, separate from the main quenching band as the maximum at 0.89 ev is separate from the main quenching band in CdS.

TABLE I. Infrared quenching data for CdS crystals.[a]

Crystal type	Relative sensitivity[b]	Percent quenching 0.89 ev	1.35 ev	1.65 ev
A[c]	530	55	96	0
B[d]	175	61	95	33
B	80	11	32	0
C[e]	54	45	81	0
A	12	63	97	99
A	11	22	48	0
B	8.5	5	14	42
C	5.5	17	30	40
A	3.7	2.5	20	81
A	2.2	35	90	97
C	1.2	2	14	80
A	1.0	7	25	76
A	0.80	4	11	31
A	0.53	7	20	35
A	0.19	0	10	75

[a] Measured with a conductivity of 5×10^{-9} mho/cm produced by the primary excitation. Dark conductivity is of the order of 10^{-10} mho/cm.
[b] Sensitivity in microamperes/100 volts for excitation by monochromator at 5150 A.
[c] Type *A* crystals are "pure" CdS.
[d] Type *B* crystals are Type *A* crystals with 10 ppm Al incorporated.
[e] Type *C* crystals are Type *A* crystals heated for 5 minutes at 900°C in sulfur vapor and quenched.

[17] In insensitive crystals, excitation is not produced by wavelengths greater than 7500 A. In sensitive crystals, the excitation spectrum may extend as far out as 9000 A.

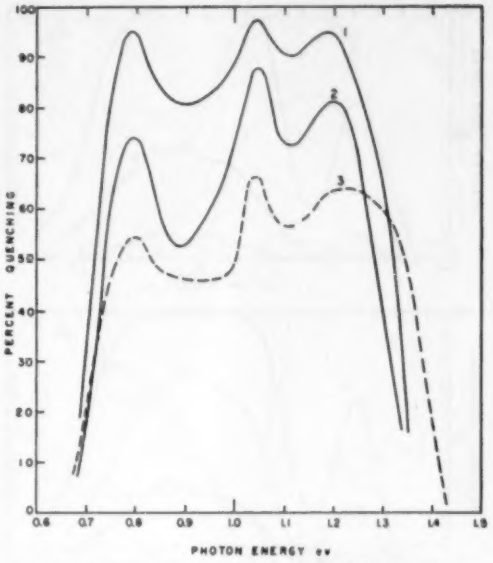


FIG. 5. Infrared quenching spectra at −189°C for three typical crystals of CdSe.

Infrared Quenching as a Function of Intensity

There are two ways to decrease the effective percent quenching: (1) decrease the intensity of the secondary radiation—as illustrated for CdS in Fig. 6(a), and (2) increase the intensity of the primary radiation—as illustrated in Fig. 6(b).

For percent quenching sufficiently far from 100% to avoid saturation effects, the magnitude of the quenching varies approximately linearly (about 0.9 power) with secondary radiation intensity, but varies quite sublinearly (about 0.7 power) with varying primary radiation intensity.

Infrared Quenching as a Function of Temperature

Infrared quenching spectra as a function of temperature, measured at the same conductivity produced by the primary radiation at each temperature, are given in Fig. 7 for a CdS crystal with nearly equal 1.35-ev and 1.65-ev peaks at room temperature, in Fig. 8 for a CdS crystal with larger 1.35-ev peak at room temperature, and in Fig. 9 for a CdS crystal with larger 1.65-ev peak at room temperature. These data show that (1) the 0.89-ev peak vanishes at temperatures below −50°C, (2) at low temperatures all curves shift maximum toward 1.65 ev, even when the room temperature curve shows only a slight quenching at 1.65 ev, (3) all quenching vanishes for temperatures above about 100°C. In analyzing the measurements of infrared quenching as a function of temperature, it must be kept in mind that the intensity of the primary excitation is varied to keep the bias conductivity constant. Insofar

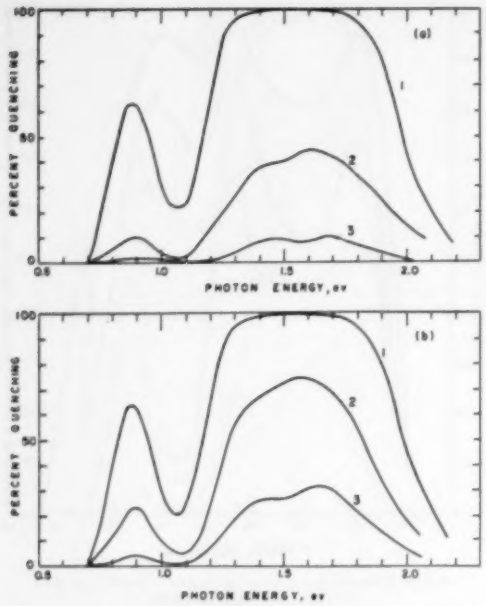


FIG. 6. Infrared quenching spectra at room temperature for a CdS crystal. (a) As a function of secondary radiation intensity: bias current produced by primary excitation of 0.07 μa/100v. (1) Full infrared quenching intensity as used elsewhere in this paper, (2) 14% of full infrared intensity, (3) 2% of full infrared intensity. (b) As a function of primary radiation intensity; secondary radiation intensity same as for (1) in (a). (1) Bias current of 0.07 μa/100v, (2) bias current of 0.36 μa/100v, (3) bias current of 1.8 μa/100v.

as the percent quenching is dependent on the ratio of primary to secondary radiation intensity, some correction for this must be introduced into the interpretation of Figs. 7 through 10. The rather slow variation of

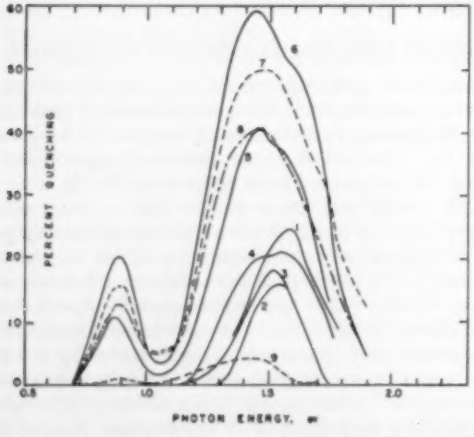


FIG. 7. Infrared quenching spectra for a CdS crystal as a function of temperature in degrees C: (1) −189, (2) −127, (3) −94, (4) −52, (5) −11, (6) 25, (7) 59, (8) 94, and (9) 133.

percent quenching with primary radiation intensity indicated in Fig. 6(b), however, shows that the major variations of percent quenching with temperature are true temperature effects.

Figure 10(a) presents data from a more detailed measurement of the temperature variation of quenching by 0.89, 1.35, and 1.65-ev photons in CdS. The inset in Fig. 10(a) shows the temperature variation of the corresponding photocurrent, measured at constant primary light intensity chosen to give a conductivity at room temperature the same as that used for measurements of infrared quenching. All quenching starts to disappear at about the same temperature that the photosensitivity decreases, i.e., about 60°C for the excitation intensity used. Curve 4 in Fig. 10(a) is the relative magnitude of excitation caused by the 1.65 ev secondary radiation,

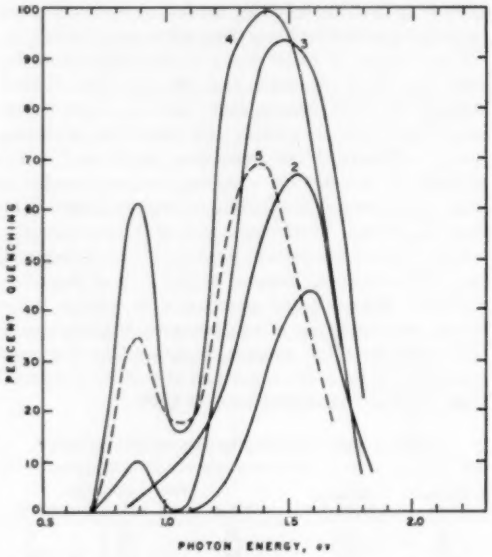


FIG. 8. Infrared quenching spectra for a CdS crystal as a function of temperature in degrees C: (1) −183, (2) −105, (3) −40, (4) 25, and (5) 86.

demonstrating that this is also temperature sensitive; the narrow temperature range over which excitation is observed corresponds to a temperature range where trapping effects are prominent.

Figure 10(b) presents detailed data on the temperature variation of quenching by 1.20, 1.05, and 0.79-ev photons in CdSe. The inset shows the variation of photocurrent at constant primary light intensity. A very rapid decrease in photosensitivity sets in at about −50°C for the excitation intensity used; for higher temperatures the infrared quenching also decreases very rapidly.

The concentration of levels in the neighborhood of the Fermi level may play a large role in determining the

magnitude of the quenching. It is evident that if the Fermi level is so located that the initation of appreciable quenching would cause it to be lowered into a region with a high concentration of filled levels, the subsequent emptying of these levels would serve both to decrease the quenching magnitude and to slow down the establishment of a steady-state condition.

Recovery from Infrared Quenching

The speed of recovery from infrared quenching is increased by the use of a higher primary radiation intensity, and is decreased by the presence of trapping effects. Figure 11 presents curves showing the recovery from infrared quenching (normalized) for two crystals of CdS at room temperature. The shape of the recovery

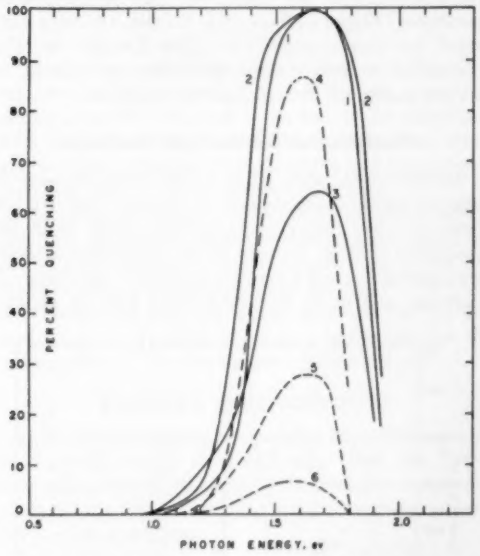


FIG. 9. Infrared quenching spectra for a CdS crystal as a function of temperature in degrees C: (1) −183, (2) −125, (3) −85, (4) −29, (5) 25, and (6) 74.

curves is exponential with time after an initial rapid recovery. There is in most crystals a faster recovery the lower the energy of the quenching photons, relatively independent of the percent quenching which is caused. Sometimes this effect is very marked, as in Fig. 11(a), whereas in other cases, the effect is less obvious although still present, as in Fig. 11(b).

Infrared Quenching and Trapping

The crystals of CdS used in this investigation show a remarkably reproducible trapping distribution from one crystal to another, as determined by measurements of thermally stimulated current. Figure 12 gives the thermally stimulated current curves for four CdS crys-

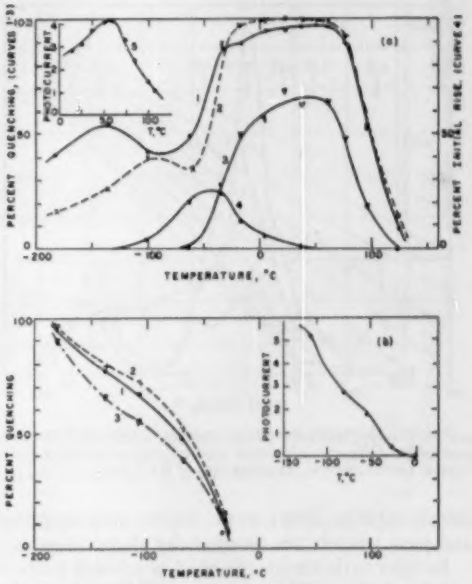


FIG. 10. (a) Temperature variation of percent quenching for a CdS crystal for photons of (1) 1.65 ev, (2) 1.35 ev, and (3) 0.89 ev. Curve (4) is the temperature dependence of the relative magnitude of excitation caused by 1.65-ev radiation. Curve (5) in the inset gives the temperature dependence of the photocurrent at constant primary radiation intensity. (b) Temperature variation of percent quenching for a CdSe crystal for photons of (1) 1.20 ev, (2) 1.05 ev, and (3) 0.79 ev. The curve in the inset gives the temperature dependence of the photocurrent at constant primary radiation intensity.

tals, all of which show evidence of a complex trapping structure involving possibly as many as seven trap depths. The solid curves are for less-sensitive pure CdS crystals, and the dotted curve, the ordinates of which have been multiplied by 0.1 before plotting, is for a sensitive CdS:Al crystal. The shift in the peaks between the dotted curve and the solid curves is exactly that which is required to give the same trap depths (as

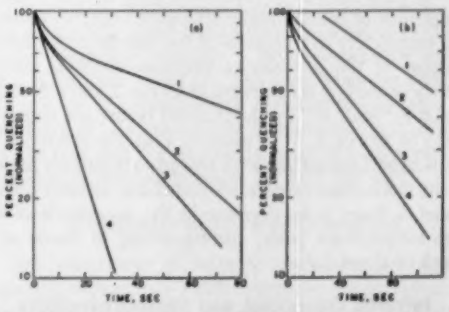


FIG. 11. Recovery from infrared quenching for two CdS crystals: (a) after quenching by radiation of wavelength (1) 7500 or 8000 A, (2) 8500 A, (3) 9000 A, and (4) 9500 A; (b) after quenching by radiation of wavelength (1) 7500 A, (2) 8500 A, (3) 10 000 A, and (4) 14 000 A.

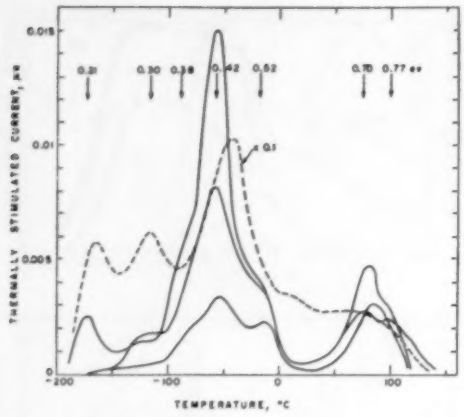


FIG. 12. Thermally stimulated current curves for three "pure" insensitive CdS crystals (solid curves) and a sensitive CdS:Al crystal (dotted curve). Heating rate of 0.77°/sec.

calculated from Fermi levels) for the peak magnitudes and peak temperature locations for all four curves.

In order to determine the effect of infrared quenching on traps, three thermally stimulated current curves were measured, as shown in Fig. 13. A curve run in the normal fashion after 15 minutes in the dark following excitation (Curve 1) is compared with a curve measured after 15 minutes in the dark following excitation, for which the sample was irradiated with 8500 A infrared between −110° and −25°C (Curve 2), and with a curve for which the sample was irradiated with 8500 A infrared for 5 minutes after 10 minutes in the dark following excitation, before the start of the thermally stimulated current measurement (Curve 3).

It must be remembered that the magnitude of the thermally stimulated current at a given temperature depends both on the concentration of traps emptying at that temperature and on the lifetime of a free electron which is dependent on the concurrent distribution of filled and empty states in the forbidden gap. An analysis of the data of Fig. 13 indicates that the decrease in magnitude of the thermally stimulated current found is due completely to a decrease in the free electron lifetime because of the infrared—a decrease which is continuously caused by the infrared in Curve 2, and a decrease which is "built in" by the infrared before the measurement of Curve 3. It is unlikely that the emptying of traps by recombination with free holes is directly related to the mechanism of infrared quenching. In both Curves 2 and 3, there is no decrease in the magnitude of the high-temperature peak, corresponding to levels at a depth of about 0.7 ev.

Infrared Quenching and Photoconductivity Rise Curves

The rise of photoconductivity in CdS crystals at room temperature depends strongly on the length of time which has elapsed since the previous excitation. After periods of darkness of several days, a pronounced *S*-shape rise curve is obtained, more than 20 seconds being required in one case for the photocurrent to rise the first one percent of its equilibrium value.

The same effect on the rise curve can be obtained by irradiating with infrared in the dark before beginning the excitation. In general the rise is slower for those infrared wavelengths which are most effective in quenching the photocurrent caused by a primary excitation.

Figure 14 shows a striking example of the type of rise curve which can be obtained. This measurement was made for a crystal of CdSe at room temperature. (The decrease in the rate of rise by irradiating with infrared in the dark before starting excitation is not found at room temperature for CdSe, but only at lower temperatures.) Figure 14 shows that the rise is initially very rapid, but then changes into a slow *S*-shape rise. The calculated electron Fermi levels for the current at several points are indicated on the curve.

Absorption of Infrared and Excitation

Measurements were made to see if any decrease in infrared transmission through a crystal of CdS could be detected when the crystal was illuminated. With the assistance of H. B. DeVore, a low intensity of infrared was passed through a crystal, which was alternately in darkness and in an intense focused spot of exciting

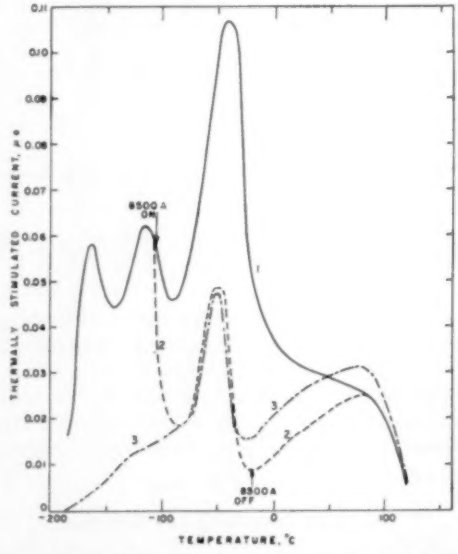


FIG. 13. Thermally stimulated current curves for a CdS crystal: (1) after 15 minutes in dark following excitation; (2) after 15 minutes in dark following excitation, exposed to 8500 A radiation between −110° and −25°C; (3) after 5 minutes exposed to 8500 A ratiation after 10 minutes in the dark following excitation. Heating rate of 0.77°/sec.

radiation. A positive effect, if present, appeared very small: about 1% decrease in transmission for 1.65-ev infrared and about 0.5% decrease in transmission for 1.35-ev infrared.

Fermi-Level Calculations

In order to give an orientation in the discussion which is to follow, Fig. 15 shows how the Fermi level varies with temperature for certain values of conductivity, given on the figure in units of mho/cm$\times 10^{-6}$. For most of the crystals tested for infrared quenching, the geometrical dimensions were such that the conductivity was numerically equal to the current for an applied voltage of 100 volts. Therefore the numbers on the curves of Fig. 15 can be interpreted approximately as the microamps of current for an applied voltage of 100 volts.

DISCUSSION

The phenomena of infrared quenching, superlinearity, temperature dependence, and slow *S*-shape growth of photocurrent in CdS and CdSe can all be correlated with the location of the calculated electron Fermi level, and hence by inference with the filling and emptying of a certain set of levels in the forbidden gap.

The main features of the following discussion, which sets forth the details of this correlation, are summarized in Table II. Important quantities to be kept in mind are the band gaps of CdS (2.4 ev) and of CdSe (1.7 ev), and the distance between the Fermi level and the conduction band in the dark at room temperature for the conductivity of the crystals used (about 0.8 ev).

Superlinear Photoconductivity

It has previously been shown[8] that superlinear photoconductivity occurs when and only when the Fermi level varies between 0.6 and 0.8 ev from the conduction band in CdS, and between 0.3 and 0.6 ev from the conduction band in CdSe.

Thermally Stimulated Current

It has also been previously shown[8] that every crystal of CdS shows a characteristic thermally stimulated current peak at 0.7 ev, and that every crystal of CdSe shows such a peak at about 0.4 ev.

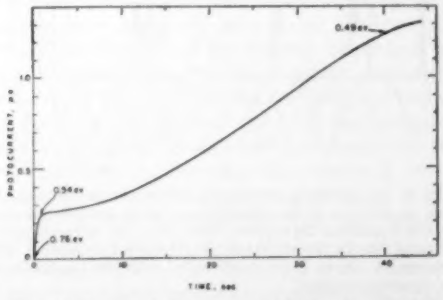


FIG. 14. Growth of photocurrent with time for a crystal of CdSe at room temperature.

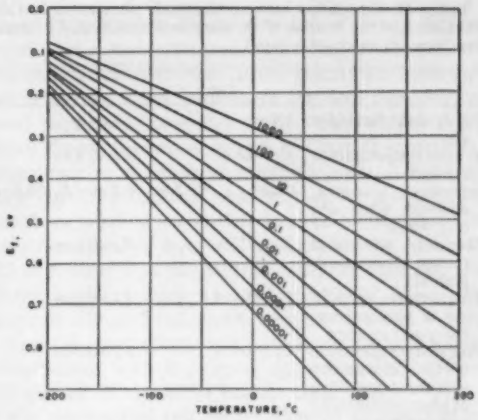


FIG. 15. Calculated distance of the electron Fermi-level from the conduction band as a function of temperature for several values of conductivity in units of mho/cm$\times 10^{-6}$.

Temperature Dependence of Photoconductivity

The photosensitivity of CdS crystals decreases very rapidly when the temperature is raised above that temperature for which the Fermi level is approximately 0.6 ev from the conduction band. The actual temperature breakpoint for photosensitivity is very sensitive to the excitation intensity, lying at lower temperatures for lower excitation intensities in such a way as to give a constant location of the Fermi level for the breakpoint. Similarly the photosensitivity of CdSe crystals decreases very rapidly when the Fermi level drops further than 0.3 ev from the conduction band with increasing temperature.[8] Because of the difference in the critical Fermi level for the two types of crystals, the breakpoint for CdS lies about 100° higher than that for CdSe.

In making many measurements with different pure CdSe crystals, occasionally a crystal is found with very low sensitivity, which has a sublinear variation of photocurrent with light intensity at room temperature instead of a superlinear variation. The correlation between the occurrence of superlinearity and the temperature dependence of photosensitivity is dramatically illustrated by the fact that the photosensitivity of such crystals is practically constant between 25° and 100°C.

Occurrence of Infrared Quenching

The magnitude of infrared quenching has a temperature dependence which is very similar to that of the photosensitivity itself. When the Fermi level drops further than 0.6 ev from the conduction band in CdS, or further than 0.3 ev in CdSe, the magnitude of the quenching decreased very rapidly; no quenching is

TABLE II. Correlation between electronic phenomena in CdS and CdSe and the location of the electron Fermi level, E_f (measured from the conduction band).

	CdS	CdSe
Width of the forbidden gap	2.4 ev	1.7 ev
At room temperature in the dark	$E_f \sim 0.8$ ev	$E_f \sim 0.8$ ev
Superlinear photoconductivity when and only when	0.6 ev $< E_f <$ 0.8 ev	0.3 ev $< E_f <$ 0.6 ev
Thermally stimulated current always shows peak at	$E_f \sim 0.7$ ev	$E_f \sim 0.4$ ev
Temperature quenching of photoconductivity sets in when	$E_f > 0.6$ ev	$E_f > 0.3$ ev
Temperature quenching of the phenomenon of infrared quenching of photoconductivity sets in when	$E_f > 0.6$ ev	$E_f > 0.3$ ev
Infrared quenching peaks at	1.65 ev 1.35 ev 0.89 ev	1.20 ev 1.05 ev 0.79 ev
Slow S-shape of photocurrent when and only when	0.6 ev $< E_f <$ 0.8 ev	0.3 ev $< E_f <$ 0.6 ev

found when the Fermi level lies further than 0.8 ev from the conduction band in CdS, or further than 0.6 ev in CdSe.

Growth of Photocurrent

Slow S-shape growth curves occur only when the Fermi level passes through that portion of the forbidden gap which is between 0.6 and 0.8 ev from the conduction band in CdS, and between 0.3 and 0.6 ev from the conduction band in CdSe. The slow growth corresponds to the time required for readjustment of empty and filled levels in the forbidden gap, holes becoming located at centers with small recombination cross section, and electrons becoming located at centers with large recombination cross section. In the case illustrated in Fig. 14, the growth is rapid as the Fermi level rises from 0.76 to 0.54 ev from the conduction band, but is then very slow as a large increase in sensitivity is caused by readjustment of empty and filled levels in the forbidden gap.

There are two simple models which are suggested by the above data. The following discussion points out the successes and difficulties of each model in accounting for the experimental results.

(1) The characteristic defect levels responsible for the phenomena of superlinearity, infrared quenching, etc. (Class II levels) lie near the conduction band.

The close correlation found in all the cases discussed above between the location of the calculated electron Fermi level and the occurrence of these photoconductivity phenomena suggest directly that the defect levels themselves are located at the distances from the conduction band indicated by the Fermi level.

The occurrence of characteristic thermally stimulated current peaks, indicating the existence of levels at these same distances from the conduction band, suggests that the current peaks are associated with the thermal emptying of the defect levels.

The quenching maxima which occur in all crystals are associated with 1.65-ev photons in CdS and 1.20-ev photons in CdSe. Assuming that these correspond to transitions from the top of the filled band (or from levels near the top of the filled band) to levels higher in the forbidden gap, the distance of these levels from the conduction band is found to be about 0.75 ev in CdS and 0.50 ev in CdSe. Thus these levels are identified with the characteristic defect levels responsible for the various effects previously described. Figure 16(a) indicates the infrared quenching transitions involved in the first model.

The 1.35-ev and 0.89-ev quenching maxima in CdS are closely related and occur prominently only in the more sensitive crystals. Although they have about the same high-temperature breakpoint as the 1.65-ev maximum, it seems probable that they are associated with a different set of defects, the presence of which by itself increases the sensitivity of the crystal by providing centers with small recombination cross section. If these defects were to have two levels associated with them, separated by 0.89 ev, the higher one lying at 1.35 ev from the filled band, then an activation energy of some 0.46 ev would be required to free a hole from the lower level, thus making it available for capture by a center with large recombination cross section. Such an activation energy would agree well with the failure to obtain quenching for 0.89-ev photons below -50°C. Such a mechanism is very similar to that proposed by Taft and Hebb.

It is interesting to note that, on the basis of studies of conductivity in CdS crystals, Kroeger[18] has concluded that the probable height of the level for an anion

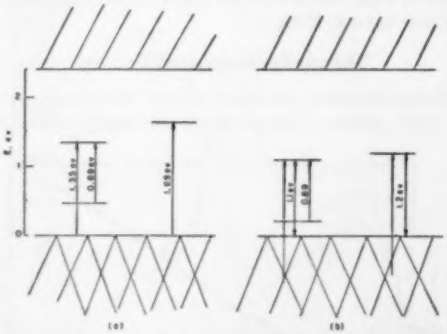


FIG. 16. (a) Infrared quenching transitions for the first model in which the location of the characteristic defect levels is associated with the location of the electron Fermi level; (b) infrared quenching transitions for the second model in which the location of the characteristic defect levels is associated with the location of the hole demarcation level.

[18] Kroeger, Vink, and van den Boomgaard, Z. physik. Chem. **203**, 1 (1954).

vacancy with one trapped electron (hence a Class I center) is about 1.8 ev above the filled band, and that the probable height of the level for a cation vacancy with one trapped hole (hence a Class II center) is about 1.3 ev above the filled band.

The most obvious failure of the model which places the characteristic defect levels near the conduction band, however, is that the simple filling and emptying of levels, lying above the Fermi level and hence empty in the dark, cannot by itself supply the phenomena of superlinearity, infrared quenching, etc., which are observed. Only the measurements of thermally stimulated current fit this model completely. As outlined in the Introduction, it is necessary for the Class II centers to be able to transfer electrons (via the conduction and valence bands) to Class I centers in the process of "sensitization" of the crystal. In the "sensitized" condition, most of the holes are located in small capture cross section Class II centers, and most of the electrons are located in large capture cross section Class I centers, thus giving few large capture cross section centers available for capturing electrons. It is clear that centers empty in the dark do not have associated with them electrons which they can transfer to Class I centers to provide this "sensitization" process.

In this model, with the defect levels located near the conduction band, the only way to obtain the phenomena of superlinearity, infrared quenching, etc. would be to have these phenomena associated with an unidentified type of levels lower in the forbidden gap, which are coupled directly with the high-lying defect levels in such a way that the occupancy of the lower levels depends on the occupancy of the upper levels.

(2) The characteristic defect levels responsible for the phenomena of superlinearity, infrared quenching etc., lie near the filled band.

This assumption is the same as was made in the Introduction and it is therefore clear that levels so located could in principle give rise to the observed phenomena.

If this model were to be adopted, it would be the location of the corresponding hole Fermi level which would be significant, instead of the actual location of the electron Fermi level. To be more precise, it would be the hole demarcation level which would be significant; we have previously pointed out the correlation between the locations of the electron Fermi level and the hole demarcation level. It would be necessary to assume that the hole demarcation level is traversing that portion of the gap near the filled band where the defect levels (Class II) are located, whenever the electron Fermi level is traversing that portion of the gap between 0.6 and 0.8 ev from the conduction band in CdS, and between 0.3 and 0.6 ev from the conduction band in CdSe. In other words, the location of the electron Fermi level is just an indication of the location of the hole demarcation level, and not of the actual location of the Class II defect levels. Since the distance of the hole demarcation level above the filled band is equal to that of the electron Fermi level below the conduction band plus a correction term depending on the ratio of the capture cross sections of the centers for holes and electrons, the exact location of the hole demarcation level requires a knowledge of this cross-section ratio. Since the capture cross section of Class II centers for holes may be as much as a million times or more as large as their capture cross section for electrons, a simple calculation shows that at room temperature the hole demarcation level will lie possibly as much as 0.3 ev further from the valence band than the electron Fermi level lies from the conduction band. When the electron Fermi level varies between 0.6 and 0.8 ev below the conduction band in CdS, therefore, the hole demarcation level is varying approximately between 0.9 and 1.1 ev above the valence band.

On the basis of this model, it must be assumed in addition that the correspondence between the thermally stimulated current peak at 0.7 ev in CdS and 0.4 ev in CdSe with the range of electron Fermi levels associated with the photoconductivity phenomena, is merely a coincidence.

Special consideration must be given the data on infrared quenching from the viewpoint of this model. Taking CdS as an example, it is found that infrared quenching bands with maxima at 1.35 ev and 1.65 ev are found. Transitions of such a large energy can be found only from low-lying levels to the conduction band, or from the filled band or low-lying levels to high-lying levels. The former transition cannot result in quenching, and may be neglected. If the data are considered evidence for actual quenching *bands* with fairly small half-width, transitions from levels near the top of the filled band to levels above the middle of the gap would be indicated, as assumed in the first model discussed above.

It is possible, however, to assume that the "band-shape" of the high-photon-energy quenching spectrum is not intrinsic to the quenching process, but is really caused by a competition between excitation and quenching by high-energy photons, excitation becoming much more prominent with increasing photon energy. On this basis, it might be argued that the significant energy in the quenching spectrum is not the energy for maximum quenching, but the lowest photon energy which will give any quenching at all. This lowest energy represents a threshold energy (the actual distance of the Class II defect levels above the filled band); for photons of higher energy than the threshold energy, the quenching arises from transitions from lower in the valence band, and is large and relatively constant. The average value for this threshold energy obtainable for the quenching data at high photon energies for CdS is about 1.1 or 1.2 ev. The most insensitive crystal, and hence the crystal with the largest value for this minimum energy, has a threshold of approximately 1.2 ev. This value does not differ too much from the estimated

possible location of the hole demarcation level for the occurrence of superlinearity, infrared quenching, etc. It would still seem likely that the 0.89-ev transition must be attributed to a transition from a level near the top of the filled band to the defect levels; in this case the level from which the 0.89-ev transition is made would be located about 0.2–0.3 ev above the filled band. Figure 16(b) indicates the infrared quenching transitions involved in this model.

The author is indebted to Dr. A. Rose, Dr. H. S. Sommers, and M. A. Lampert for many helpful discussions.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Frequency Dependence of Magnetic Resonance in α-Fe_2O_3

HIROO KUMAGAI, HIDETARO ABE, KAZUO ÔNO, IZUO HAYASHI, JUNJI SHIMADA, AND KENZO IWANAGA
Institute of Science and Technology, University of Tokyo, Tokyo, Japan

(Received April 12, 1955)

Magnetic resonance absorption in single crystals of natural α-Fe_2O_3 was examined in the wavelength range from 5.4 mm to 6.2 cm. Resonance fields H_r observed in the basal plane do not give a quadratic relation with the normal resonance field H_0 as predicted by the usual Kittel formula, but rather the following linear relationship: $h\nu/g\beta = H_0 = H_r + H_A$, where H_A is 4.2_4K oe and g is 2.5_4.

1. INTRODUCTION

MAGNETIC resonance in single crystals of α-Fe_2O_3 has been observed at 1.25-cm wavelength by Anderson *et al.*[1] We refer to their paper as (A) in the following. They assumed the usual Kittel formula[2]:

$$H_0 = [H_r(H_A + H_r)]^{\frac{1}{2}}, \quad (1)$$

where H_0 is the normal resonance field, H_r the observed resonance field in the basal plane, and H_A the anisotropy field in the direction of the ternary axis. Then, under the assumption $g = 2$, they obtained the anisotropy field $H_A \simeq 30\,000$ oe. These resonance absorptions may be caused by the parasitic ferromagnetism of this crystal.

We have examined single crystals of natural α-Fe_2O_3 (α-hematite) by the usual method of magnetic resonance absorption at seven different wavelengths (from 5.4 mm to 6.23 cm). The resonances were observed at various orientations of static magnetic field and at room temperature except where the temperature is shown.

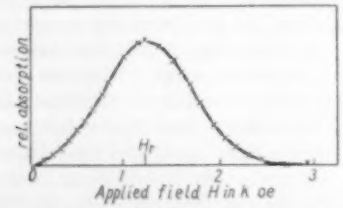


FIG. 1. Resonance absorption in α-Fe_2O_3 at $\lambda = 15.3$ mm and at room temperature. The static magnetic field is applied in the basal plane, i.e., perpendicular to the ternary axis.

[1] Anderson, Merritt, Remeika, and Yager, Phys. Rev. **93**, 717 (1954).
[2] C. Kittel, Phys. Rev. **73**, 155 (1948).

2. EXPERIMENTAL

The samples used were obtained from Sagashima, in the western part of Japan. The crystals have metallic luster and are disk-shaped with a diameter of about 10 mm and a thickness of 0.5–1 mm. They have (magnetically) hexagonal symmetry. The symmetry axis is perpendicular to the disk plane, so we call this the "basal plane." They easily cleave in this plane. Since the absorption is very strong and the nonmagnetic loss is fairly large, thin specimens of about 5-mm² area were used. At shorter wavelengths, specimens of only about 0.5-mm² area were used.

The method of measurement is almost the same as described elsewhere,[3,4] but in the present case it differs from the past descriptions in two respects. One is that the field modulation method was used to obtain line widths, though the direct galvanometer method was also used at some wavelengths. The other is that we omitted the procedure of tuning the sample cavity to cancel out the dispersion effect accompanied by reso-

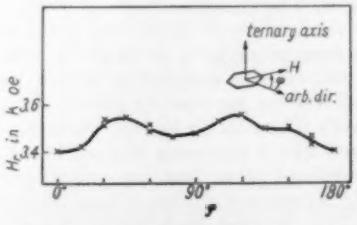


FIG. 2. Variation of resonance field H_r in the basal plane. ($\lambda = 11$ mm.)

[3] H. Kumagai *et al.*, J. Phys. Soc. Japan **9**, 369 (1954).
[4] H. Abe *et al.*, J. Phys. Soc. Japan **9**, 814 (1954).

nance. So the "width" in this article may be larger than the true width of the absorption line.

3. RESULTS AND DISCUSSIONS

When the static magnetic field is in any direction in the basal plane, a single peak is observed as shown in Fig. 1. When the static field is rotated in this plane, the resonance field varies slightly as in Fig. 2. We cannot clearly find the regular dependence with φ,

$$H_r = \text{const} + A \sin 6\varphi, \tag{2}$$

as described in (A). Even if there existed this relation, A seems to be smaller than 70 oe in the case of both $\lambda = 11$ mm and 5.4 mm.

The observed line widths are slightly larger than those of (A), as shown in Table I. Since, however, in our present experiments we are not principally interested in the line width, some precautions are omitted which must be taken for the measurement of line width. So the difference in the observed widths in Table I has little physical meaning.

The observed resonance field H'_{res} in the plane containing the ternary axis is similar to that in (A). It satisfies approximately, at $\lambda = 6.5$ mm and 11 mm,

TABLE I. Observed line widths of resonance absorption in α-Fe_2O_3.

λ in mm	Observed width in oe	Reference
5.4	930	
11	1000	
12.5	600	(A)

a relation

$$H'_{res} = H_r / \sin\theta, \tag{3}$$

as shown in Fig. 3, where θ is the angle between the ternary axis and the direction of the static field, and H_r is the resonance field in the basal plane. Equation (3) shows that the magnetization responsible for this resonance absorption is confined within the basal plane, which fact might be ascribed to an axial anisotropy field. This is compatible with Néel's hypothesis[5] given in explanation of the magnetostatic behavior of this crystal.[6]

At 3-cm wavelength, the resonance occurs at about zero field both when the static field is parallel to and when it is perpendicular to the ternary axis, and the absorption intensity is half its maximum (that at zero field) at 1100 oe and 340 oe respectively. The absorption coefficient in this case has a hysteresis relative to the applied static field. Even at $\lambda = 6.2_3$ cm, we can see a weak tail of absorption at zero field. The maximum absorption at this frequency is only about 0.5% of the incident power which is much weaker than that at shorter wavelengths.

[5] L. Néel, Revs. Modern Phys. **25**, 59 (1953).
[6] L. Néel and R. Pauthenet, Compt. rend. **234**, 2172 (1952).

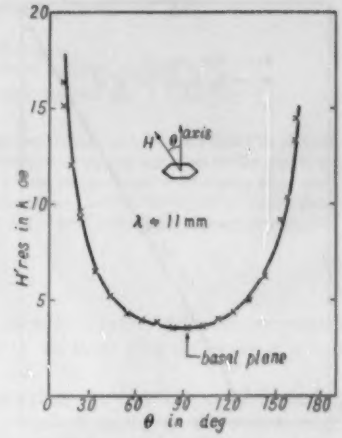


FIG. 3. Variation of resonance field H'_{res}, when the static field is rotated in a plane containing the ternary axis.

To obtain the resonance field H_r in the basal plane, we choose the minimum of the resonance field as obtained in experiments such as those shown in Fig. 2, because errors in setting the specimen must bring about higher resonance fields as shown in Fig. 3.

Data on the resonance field thus obtained are listed in Table II. They are also plotted in Fig. 4 as a function of measuring frequency ν, in which the plotted circles are much larger than the range of observed values. In measurements at 3 and 6 cm wavelengths, the static magnetization of the sample may not be saturated, as the static magnetic field is low. At other shorter wavelengths, the static magnetization of the sample may be considered to be at saturation. When we omit the results at longer wavelengths where the static magnetization is not saturated, we obtain the following linear relation between the resonance field H_r and the normal one H_0,

$$H_0 = H_r + H_A, \tag{4}$$

where

$$H_0 = h\nu/g\beta. \tag{5}$$

Our experimental data are best fitted to the above relation (4) when we take the values $g = 2.5_4$ and $H_A = 4.2_8$ kilo-oersteds.

Since a strong anisotropy field really exists in the

TABLE II. Frequency dependence of resonance field H_r in the basal plane of α-Fe_2O_3 natural crystal.

λ in cm^{-1}	H_r in oe
1.84	11 380±100
1.55	8800±100
1.21_6	5980±100
0.90_8	3400± 50
0.65_3	1300± 70
0.31_8	~0
0.16	...

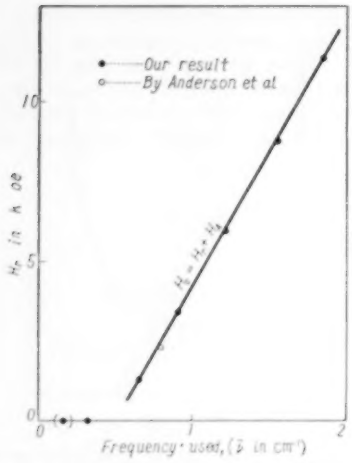


FIG. 4. Frequency dependence of the resonance field H_r in the basal plane.

direction of the ternary axis, which can be deduced from the results shown by Eq. (3) at this temperature, Kittel's considerations on ferromagnetic resonance offer a quadratic relation, Eq. (1), between the observed resonance field H_r and the normal one H_0. In his results, a linear relation can be obtained only when the static field is applied in the axial direction of the anisotropy field in uniaxial crystals.

We have plotted, in Fig. 5, the value

$$x=[H_r(H_A+H_r)]^{\frac{1}{2}}/H_0$$

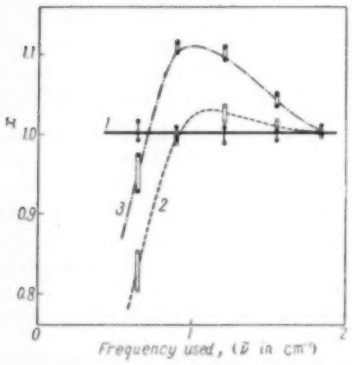


FIG. 5. Quantity x normalized at the highest frequency, where x is defined as follows. Case (1): $x=(H_r+H_A)/H_0$ with $H_A=4.2_5$ kilo-oersteds. Case (2): $x=[H_r(H_A+H_r)]^{\frac{1}{2}}/H_0$ with $H_A=30$ kilo-oersteds. Case (3): $x=(H_r)^{\frac{1}{2}}/H_0$. The quantity x should be independent of $\bar{\nu}$.

in Eq. (1) as a function of ν; x should be a constant for a suitable value of H_A when Eq. (1) is valid. We cannot realize a constant value of x with any value of H_A. On the other hand, $x=(H_r+H_A)/H_0$ maintains a constant value for $H_A=4.2_5$ kilo-oersteds. This shows that only the linear relation of Eq. (4) holds within experimental error.

The result in (A) can be fitted to our linear relation (4) as shown in Fig. 4.

When the sample is cooled the absorption intensity observed at $\lambda=6.5$ mm does not change from room temperature to $-10°$C and does not shift or broaden appreciably; but the absorption intensity observed at $\lambda=3$ cm slowly decreases and goes down to zero at about $-50°$C as shown in Fig. 6.

Our transition temperature is lower than that of the static susceptibility χ observed by Morin,[7] that of neutron diffraction observed by Corliss *et al.*,[8] and that of resonance absorption in a synthetic crystal observed by Anderson *et al.*[1] This difference may be due to an

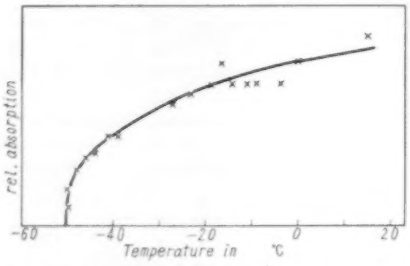


FIG. 6. Temperature dependence of absorption height with H in the basal plane at $\lambda \doteqdot 3$ cm.

impurity (e.g., TiO_2), but it cannot be considered to affect the essential features of the resonance, expressed by Eq. (4).

It is desirable to repeat the above experiments with synthetic crystals or natural ones from other sources, but, so far, they are not available to us.

4. ACKNOWLEDGMENTS

We wish to express our sincere thanks to Dr. P. W. Anderson (Bell Telephone Laboratories) who suggested this work during his stay in Japan as a visiting professor at the University of Tokyo. Our thanks are also due to Professor T. Itoh (Department of Mineralogy, University of Tokyo) for his kindness in offering us samples of α-Fe_2O_3.

[7] R. J. Morin, Phys. Rev. **78**, 819 (1950).
[8] Corliss, Hastings, and Goldman, Phys. Rev. **93**, 893 (1954).

Atomic Heat of Graphite between 1 and 20°K†

P. H. KEESOM AND N. PEARLMAN
Department of Physics, Purdue University, West Lafayette, Indiana
(Received December 27, 1954; revised manuscript received May 4, 1955)

The atomic heat of pure artificial polycrystalline graphite has been measured between 1 and 4°K and 10 and 20°K. Precautions were taken to ensure that no error was introduced by gas adsorbed on the graphite. Below 2°K the atomic heat can be represented as the sum of a term proportional to T^3 arising from low-frequency lattice waves and a term proportional to T due to electrons: $C=0.0325T^3+0.031T$ millijoules/mole degree. Between 2.25°K and 4.5°K, $C=0.1157^3+0.031T-0.237$ millijoules/mole degree; between 10 20°K, $C=0.2087^3-6.8$ millijoules/mole degree.

I. INTRODUCTION

CONSIDERABLE attention has been given recently to the band structure of graphite[1–3] and to the vibration spectrum of its lattice.[3–10] This lattice is highly anisotropic, being made up of planes of atoms strongly bound together in a hexagonal network, the planes being held together by weak forces of the van der Waals type. The bond length in the planes is 1.42 A, while the distance between planes is 3.37 A. As a consequence, the Brillouin zone is foreshortened in the direction corresponding to wave propagation perpendicular to the planes. The lattice contribution to the atomic heat, C_L, is therefore found[3–6,8,9] to be proportional to T^2 at low temperatures, rather than to T^3, as is the case for ordinary solids. There is disagreement, however, as to whether the T^2 dependence will hold down to 0°K,[3–5] or whether below a few degrees absolute, a T^3 region will be found.[6,8,10] This question has not been settled by the recent measurements of de Sorbo and Tyler,[11] which show that the atomic heat is proportional to T^2 between 13°K (their lowest temperature) and 54°K. One of the purposes of our atomic heat measurements down to 1°K was thus to investigate the temperature dependence of C_L at very low temperatures. Shortly after a summary of this paper was originally submitted for publication, values were reported for the atomic heat of graphite from 1.5 to 90°K.[12] These data agree very well with those reported here in the over-lapping temperature ranges, but they are interpreted by Bergenlid *et al.*, in a way which leads to a rather different temperature dependence than we find. This difference is discussed below in Sec. IV.

In addition, we were interested in finding the electronic contribution, C_E, to the atomic heat. This was calculated by Komatsu and Nagamiya[3] from the band structure described by Wallace,[1] who found a band that is full at 0°K and an empty higher band which just touches the former. Their rough calculation showed that C_E and C_L would be comparable in size at 1°K.

II. EXPERIMENT

The sample was a cylinder of pure artificial graphite, about 4 cm in diameter and weighing about 170 grams, obtained from the Stackpole Carbon Company. Cigarette paper was fastened to the cylinder with glyptal lacquer, and the thermometer wires (phosphor-bronze in the helium range, lead in the hydrogen range) wound thereon and also secured with glyptal. The thermometers were calibrated against the vapor pressures of the respective baths, using the 1948 scale[13] for helium and the NBS formulas[14] for hydrogen. Insulated constantan wire wrapped on the cylinder was used for the heater. The sample was hung in a vacuum can, and the measurements carried out in the manner described previously.[15] The heat capacity of the glyptal, wires, etc., was less than one percent of the total.

Since graphite is an excellent adsorbent, we anticipated that some of the exchange gas introduced into the vacuum can prior to calibration of the thermometers would remain adsorbed on the sample after most of it was pumped away following calibration. Such adsorbed gas would introduce systematic errors in the heat capacity determinations. To estimate the magnitude of these errors, we carried out two preliminary runs. In run *A*, the exchange gas was pumped away for several hours while the sample was being heated and thereby kept at about 5°K. When the system pressure was constant, heating was stopped and the sample came to bath temperature (4.2°K) in less than an hour. The heat capacity was then measured at 4°K and below.

† Work supported by Signal Corps contract, and U. S. Atomic Energy Commission contract.

[1] P. R. Wallace, Phys. Rev. **71**, 622 (1947).
[2] C. A. Coulson and R. Taylor, Proc. Phys. Soc. (London) **A65**, 815 (1952).
[3] K. Komatsu and T. Nagamiya, J. Phys. Soc. Japan **6**, 438 (1951).
[4] V. V. Tarassov, Compt. rend. acad. sci. (U.R.S.S.) **46**, 20, 110 (1945).
[5] R. W. Gurney, Phys. Rev. **88**, 465 (1952).
[6] P. G. Klemens, Australian J. Phys. **6**, 405 (1953).
[7] J. P. Hobson and W. A. Nierenberg, Phys. Rev. **89**, 662 (1953).
[8] J. Krumhansl and H. Brooks, J. Chem. Phys. **21**, 1663 (1953).
[9] H. B. Rosenstock, J. Chem. Phys. **21**, 2064 (1953).
[10] T. Nagamiya and K. Komatsu, J. Chem. Phys. **22**, 1457 (1954).
[11] W. de Sorbo and W. W. Tyler, J. Chem. Phys. **21**, 1660 (1953).
[12] Bergenlid, Hill, Webb, and Wilks, Phil. Mag. **45**, 851 (1954).
[13] H. van Dijk and D. Shoenberg, Nature **164**, 151 (1949).
[14] Wolley, Scott, and Brickwedde, J. Research Natl. Bur. Standards **41**, 379, 1948.
[15] N. Pearlman and P. H. Keesom, Phys. Rev. **88**, 398 (1952).

TABLE I. Atomic heat of graphite.

n	T °K	ΔT deg	C millijoules / mole degree
2	4.392	0.04862	2.049
3	4.391	0.04839	2.059
4	2.990	0.06667	0.891
5	3.023	0.06357	0.935
1	4.257	0.0901	1.982
2	4.333	0.06173	2.079
3	4.366	0.06151	2.087
4	3.995	0.07320	1.753
5	4.045	0.07120	1.802
6	3.755	0.04650	1.467
7	3.793	0.04469	1.526
8	3.406	0.05632	1.210
9	3.445	0.05522	1.234
10	3.092	0.07264	0.9376
11	3.147	0.07021	0.9702
12	2.748	0.07150	0.7168
13	2.805	0.06829	0.7506
14	2.578	0.08606	0.5954
15	2.655	0.07928	0.6464
16	2.200	0.07267	0.3997
17	2.236	0.07350	0.3951
18	1.778	0.07746	0.2420
19	1.838	0.06769	0.2769
20	1.861	0.07199	0.2604
21	1.632	0.09582	0.1956
22	1.681	0.09118	0.2056
23	1.235	0.07909	0.1019
24	1.255	0.07741	0.1041
25	1.246	0.07584	0.1062
26	1.063	0.07868	0.0715
27	1.081	0.07891	0.0713
1	9.20	0.07631	10.83
2	9.24	0.07328	11.27
3	9.27	0.2016	11.09
4	8.96	0.2245	9.96
5	9.14	0.2124	10.52
6	9.52	0.1890	11.84
7	9.70	0.1752	12.87
8	9.89	0.1701	13.14
9	10.48	0.1414	15.88
10	10.65	0.1346	16.64
11	11.06	0.1203	18.63
12	11.18	0.1170	19.14
13	11.74	0.1001	22.39
14	11.89	0.1935	23.17
15	12.08	0.1850	24.24
16	12.80	0.1622	27.68
17	12.97	0.1569	28.62
18	13.10	0.0736	30.53
19	14.00	0.1314	34.16
20	14.12	0.1284	34.97
21	14.16	0.1260	35.64
22	14.94	0.1141	39.40
23	15.05	0.1126	39.92
24	16.42	0.0911	49.45
25	16.50	0.0901	50.02
26	17.36	0.0811	55.65
27	17.43	0.0801	56.36
28	18.34	0.0722	62.58
29	18.42	0.0720	62.77
30	19.33	0.0661	68.47
31	19.40	0.0655	69.15
32	20.30	0.0602	75.20
33	20.36	0.0593	76.49

In run B, the sample was heated in steps up to about 10°K, and the system pressure became constant only after about seven hours of pumping. Heating was then stopped, and the sample did not reach bath temperature until about eight hours later. The heat capacity was measured at several temperatures, and found to be about 10% lower than the values at corresponding temperatures measured during run A. The results reported below were obtained during a third run for which the pumping procedure of run B was followed. In this run and also run B, increasing the sample temperature when the system pressure had finally become constant after prolonged pumping produced no increase in system pressure, thus indicating that substantially all of the exchange gas had been pumped away.

The possibility that air adsorbed on the sample at room temperature might contribute significantly to the measured heat capacity was also investigated. The sample was heated to about 300°C and the gas driven off, less than one cc (NTP), was collected. The corresponding heat capacity below 20°K would be less than 0.1% of the total, as estimated from the data of Morrison and Szasz[16] for N_2 adsorbed on TiO_2.

III. RESULTS

The results of our measurements in the helium range and the hydrogen range are presented in Table I. They are given in graphical form in several different ways in Figs. 1–3. In Fig. 1, we plot C/T *versus* T^2 (left-hand ordinates) and C *versus* T^2 (right-hand ordinates) for the helium range data. Here C is the total measured atomic heat. We also plot in Fig. 1, C_L *versus* T^2, where C_L is the lattice contribution to the atomic heat,

$$C_L = C - C_E, \tag{1}$$

where C_E is the electronic contribution. This is estimated on the assumption that the straight line obtained below 2°K in the C/T *versus* T^2 plot results from the sum of a term proportional to T^3 (C_L) and a linear term (C_E) whose coefficient is given by the intercept of the line. The equation for the line is

$$C = 0.0325T^3 + 0.031T \text{ millijoule/mole degree.} \tag{2}$$

The data evidently do not conform to this equation above about 2°K. In the range between 2.25° and the boiling point, the C *versus* T^2 appears to be a good straight line, with the equation

$$C = 0.1187T^2 - 0.177 \text{ millijoule/mole degree.} \tag{3}$$

The straight line fit is improved if C_E is subtracted, resulting in the line labeled C_L, which has the equation

$$C_L = 0.1157T^2 - 0.237 \text{ millijoule/mole degree.} \tag{4}$$

The data between 10 and 20°K are plotted in Fig. 2, in the form C *versus* T^2. A fairly satisfactory fit is found with the line

$$C = 0.208T^2 - 6.8 \text{ millijoules/mole degree.} \tag{5}$$

[16] J. A. Morrison and G. J. Szasz, J. Chem. Phys. **16**, 280 (1948).

The data of de Sorbo and Tyler[11] are also plotted in Fig. 2. These data are about 10% higher; the dashed line is that given by de Sorbo and Tyler[11] as fitting their data between 13 and 54°K:

$$C = 0.216T^3 \text{ millijoule/mole degree.} \quad (6)$$

The source of the discrepancy between our results and those of de Sorbo and Tyler[11] is not clear. Their material was in the form of powder, and was contained in a copper calorimeter during measurement. It is unlikely that the difference is a particle size effect, since they give the average thickness of the crystallite plates as 230 A, while x-ray measurements carried out on our sample by Mr. W. D. Derbyshire of this laboratory, using the broadening of the 002 line, give about 300 A. The values reported by Bergenlid *et al.*,[12] are also lower than those of de Sorbo and Tyler[11] below about 25°K.

We will discuss below the analysis of our data in terms of Figs. 1 and 2. In these graphs, the straight lines are drawn solid or dashed where they pass through the points on which they are based, and dotted where they are extrapolated beyond these points. The solid or dashed lines do not deviate from their points by more than the experimental error (about three percent). In order to see whether any single equation of the form $C = aT^n$, [such as Eq. (6) found by de Sorbo and Tyler[11]] would fit our data, we plotted log C *versus* log T in Fig. 3. The straight line drawn there, which has the equation

$$C = 0.0605T^{2.33} \text{ millijoule/mole degree,} \quad (7)$$

appears to be a reasonably good fit. However, it deviates from the experimental points by as much as 10% at several temperatures, and so is not to be preferred to the lines in Figs. 1 and 2.

Bergenlid *et al.*[12] presented their data in a log-log plot, but they point out that a single straight line will not fit adequately over their entire temperature range. The smooth curve they draw through the points has a slope of 1.8 at 90°K; the slope increases to 2.4 at 4°K. At their lowest temperatures their points fall near a line with the equation

$$C = 0.067T^{2.4} \text{ millijoule/mole degree,} \quad (8)$$

which agrees with Eq. (7) to within the combined accuracies. The same systematic deviations from the line apparent in Fig. 3 occur as well with respect to the line in their graph.

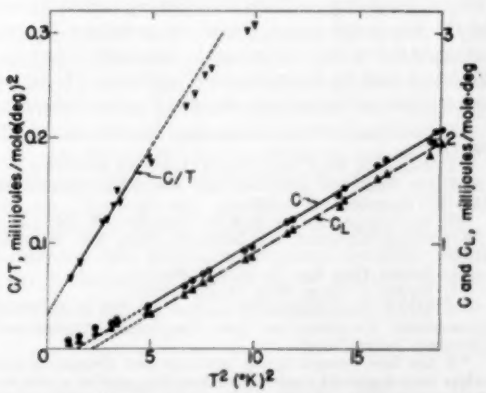


FIG. 1. Atomic heat of graphite below 4°K: ▼ C/T *versus* T^2 (left-hand ordinates); ● C *versus* T^3; ▲ C_L (total atomic heat minus electronic contribution) *versus* T^3 (both on right-hand ordinates).

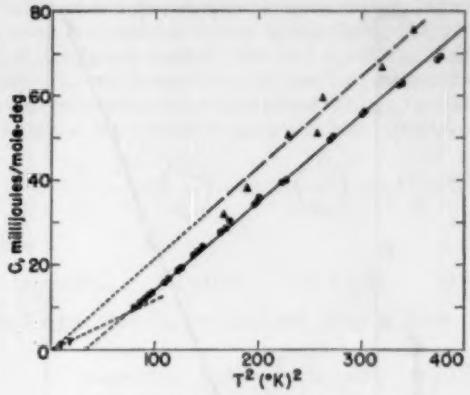


FIG. 2. Atomic heat of graphite *versus* T^3: ▲ de Sorbo and Tyler; ● our data.

IV. DISCUSSION

The temperature dependence to be expected at very low temperatures for C_V of graphite depends on the assumptions made concerning: (a) the temperature dependence of C_L; (b) the presence or absence of a significant contribution to C_V besides C_L, such as an electronic contribution, C_E.

(a) As was mentioned in the Introduction, different theoretical treatments lead to either quadratic or cubic dependence of C_L on T, and it is of course possible that the true temperature dependence differs from both of these. A possible source of evidence to resolve this question might be found in measurements of the thermal conductivity, κ, at low temperatures. It can be shown[17] that

$$\kappa \propto v C_V \lambda,$$

where v is the velocity, C_V the heat capacity per unit volume, and λ the mean free path of the "carriers" of thermal excitation. In graphite heat will be carried mainly by phonons[18] and at low temperatures λ will be limited by the size of the crystallites,[18] and so will not depend on temperature. Since v is also essentially

[17] P. Debye, *Vortrage uber die Kinetische Theorie der Materie und der Elektrizität* (B. G. Teubner, Leipzig, 1914), pp. 17–60.

[18] R. Berman, Phil. Mag. Supplement **2**, 103 (1953); Proc. Phys. Soc. (London) **A65**, 1029 (1952).

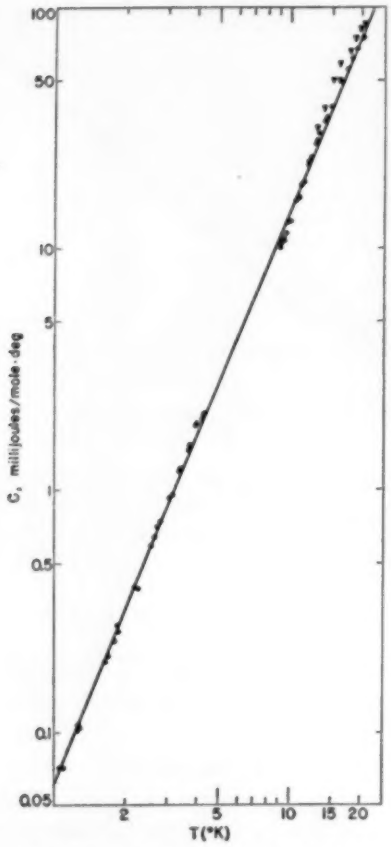


FIG. 3. Atomic heat of graphite below 20°K: ▼ de Sorbo and Tyler; ● our data.

independent of temperature, κ will be proportional to C_L. Hence the results of Berman[18] which show that in graphite κ varies as T^n with n between 2.5 and 2.7 for T between 3 and 40°K, could be interpreted as indicating a similar T dependence for C_L. Klemens,[6] however, attempted to show that this behavior could be reconciled with T^2 variation of C_L by considering the different contributions to κ from the longitudinal and transverse phonons. On the other hand, Smith has suggested[19] that the deviation from T^2 dependence observed by Berman is due to intergranular non-graphitic material in the polycrystalline samples he measured. Smith has recently found[20] that κ is proportional to T^2 for T between 4 and 70°K in samples of natural graphite, thus lending support to this point of view. To sum up, there appears to be no independent evidence concerning the temperature dependence of C_L below 2°K; T^2 dependence would seem likely, at least above 4°K, from Smith's thermal conductivity measurements.

(b) The calculations of Komatsu and Nagamiya,[3] based on Wallace's theory of the band structure of graphite, indicate that in the neighborhood of 1°K the electronic heat capacity might be comparable to that of the lattice. Attempts have been made to correlate the results of measurements[21–25] of the electrical resistivity of graphite with Wallace's theory, with inconclusive results. The theory applies to perfect crystals, while the material available is either polycrystalline artificial graphite or natural crystals of doubtful purity. Attempts to modify it[22,26,27] in order to take these factors into account have not succeeded in improving the agreement with experimental results. On the other hand, Lothar Meyer has reported[28] that new measurements on samples first investigated several years earlier gave results very different from those found originally. He has also found inconsistencies in results on apparently similar samples. This indicates that the discrepancies between Wallace's theory and experimental resistivity values may be due to difficulties in controlling extraneous conditions to which the samples may be very sensitive, such as thermal history, traces of impurities, etc., rather than to defects in the theory itself.[29] Hence these discrepancies between Wallace's theory and the electrical measurements do not in themselves argue strongly against the existence of an electronic contribution to the heat capacity.

On the basis of these considerations, it appears reasonable to conclude from our data that the lattice atomic heat of graphite is proportional to T^3 below 2°K and that there is an additional contribution from the electrons, which is about equal to that of the lattice at 1°K. The electronic term is about three times larger than that calculated by Komatsu and Nagamiya [see Eq. (31) of reference 3]. One of the factors entering into the calculation is the value of the resonance-energy integral involving electrons in adjoining carbon atoms in the hexagonal plane. From the measured C_E we calculate 0.9 ev for this integral, compared with 3 ev, the value used by Komatsu and Nagamiya. The latter value is derived from x-ray emission line fine structure.

[18] A. W. Smith, Phys. Rev. **93**, 952 (1954).
[19] A. W. Smith, Phys. Rev. **95**, 1095 (1954).
[21] A. Goetz and A. Holser, Trans. Am. Electrochem. Soc. **82**, 391 (1942).
[22] S. Mrozowski, Phys. Rev. **77**, 838 (1950); **85**, 609 (1952).
[23] Meyer, Picus, and Johnston, Natl. Bur. Standards Circular 519, 1952 (unpublished), p. 249.
[24] W. W. Tyler and A. C. Wilson, Jr., Phys. Rev. **89**, 870 (1953).
[25] Reynolds, Hemstreet, and Leinhardt, Phys. Rev. **91**, 1152 (1953).
[26] D. Bowen, Phys. Rev. **76**, 1878 (1949).
[27] G. Hennig, J. Chem. Phys. **19**, 922 (1951).
[28] L. Meyer, paper presented December 21, 1953 at the Third International Conference on Low Temperature Physics and Chemistry (unpublished).
[29] It has been brought to our attention that Krumhansl and others have suggested screw dislocations as a possible source for the irreproducibility of conductivity results. These would provide a mechanism for conduction perpendicular to the hexagonal planes without requiring transistors between planes. Unfortunately, no information seems to be available on screw dislocations in graphite.

Coulson[2] has pointed out, however, that the calculation is very sensitive to the shape of the electron level-density curve at the band edge. Since this is not known with great accuracy on the basis of Wallace's theory, it is more likely that this factor, rather than the resonance-energy integral is the source of the discrepancy between measured and calculated values of C_E, and the lack of quantitative agreement between them is not surprising.

Our results are thus in at least formal agreement with any theory which predicts that C_L is proportional to T^3 below a certain transition temperature T_0 and to T^2 at somewhat higher temperatures. Quantitative agreement with any particular theory is hardly to be expected, however, because of the lack of sufficient information about the interatomic force constants which enter.[9,30] It is nevertheless interesting to compare our results with the analysis of Krumhansl and Brooks,[6] which is the most detailed of the treatments of this type.

There are various ways of describing the normal mode atomic displacements in graphite.[3,5,6] Krumhansl and Brooks use the procedure suggested by Lonsdale,[31] according to which the polarization vector of the displacement wave is taken either perpendicular to the hexagonal planes (bending modes; z modes) or in the plane (stretching modes; xy modes). The relatively weak forces between the planes are largely associated with the z modes, while the much stronger intraplanar forces are largely associated with the xy modes. For this reason the z modes will be more strongly excited at low temperatures and hence be the major source of the heat capacity. Therefore Krumhansl and Brooks discuss only these modes in detail.

Using only nearest-neighbor forces, they find that the constant energy surfaces for the z modes in the Brillouin zone are prolate ellipsoids for frequencies much less than

$$2\pi\nu_{1z}=(2S_1/m)^{\frac{1}{2}}, \tag{9}$$

where m is the mass of the carbon atom and S_1 is the restoring force constant between nearest neighbors in adjacent planes. Thus for $\nu\ll\nu_{1z}$ the vibration spectrum for the z modes, $g_z(\nu)$, is proportional to ν^2 as in the Debye continuum approximation, which is known to apply also for real crystals at sufficiently low frequency. As ν approaches ν_{1z} the constant frequency surfaces become flattened near the top and bottom of the Brillouin zone. For $\nu>\nu_{1z}$ they are truncated ellipsoids. Krumhansl and Brooks approximate the behavior of $g_z(\nu)$ in the neighborhood of ν_{1z} by setting it proportional to ν^2 up to ν_{0z} and proportional to ν above ν_{0z}. From the relation

$$\int_0^{\nu_{0z}}\text{const}\times\nu^2d\nu=\int_0^{\nu_{1z}}g_z(\nu)d\nu=n_zN_0, \tag{10}$$

where $g_z(\nu)$ is their exact analytical expression, (which however is impractical for use in calculating C_L), and N_0 is Avogadro's number, they find $\nu_{0z}=0.9\nu_{1z}$. Then below T_{0z}, the temperature of transition from cubic to quadratic temperature dependence of $C_{L(z)}$, the contribution of the z modes to the lattice heat capacity,

$$\frac{C_{L(z)}}{R}=\frac{8}{3\pi}\cdot\frac{12\pi^4}{15}n_z\left(\frac{T}{\theta_{1z}}\right)^3=66.2n_z\left(\frac{T}{\theta_{1z}}\right)^3=a_3T^3, \tag{11}$$

where

$$h\nu_{1z}=k\theta_{1z},\qquad T_{0z}\approx0.3\theta_{0z};\qquad \theta_{0z}\approx0.9\theta_{1z}. \tag{12}$$

For T larger than T_{0z} but much less than θ_{mz}, where

$$h\nu_{mz}=k\theta_{mz};\qquad N_0=\int_0^{\nu_{mz}}g_z(\nu)d\nu, \tag{13}$$

they find

$$C_{L(z)}/R=14.4(T/\theta_{mz})^2-n_z=a_2T^2-a_0. \tag{14}$$

If we interpret our results in terms of these equations, then T_{0z} is about 2°K and θ_{0z} is 6–7°K. In order to calculate θ_{mz} and n_z, some assumption must be made concerning $C_{L(xy)}$ at the lowest temperatures. The value of θ_{mxy} is found by Krumhansl and Brooks from the atomic heat at high temperatures to be about 2500°K. If $C_{L(xy)}$ is proportional to T^2 at the lowest temperatures, it would be equal to $4.61\times10^{-6}T^2R$, since there are twice as many xy modes as z modes. This is one-third the total C_v at 2°K, and almost two-thirds at 1°K, but we observe no such quadratic term between 1 and 2°K. Thus we conclude that $C_{L(xy)}$ will be given by an expression like Eq. (11) below T_{0xy} and like Eq. (14) above T_{0xy}, where T_{0xy} is not less than 2°K. If we assume that the xy mode equal energy surfaces in the Brillouin zone are similar to those of the z modes[32] then there will be a quadratic-linear transition in the vibration spectrum of the xy modes as well, at a frequency

$$\nu_{1xy}\approx\nu_{1z}v_{xy}/v_z, \tag{15}$$

where v_{xy} and v_z are the respective wave velocities for the two types of mode. Since v_{xy} would be expected to be larger than v_z, we would expect T_{0xy} to be larger than T_{0z}. If the velocity ratio in Eq. (15) were small enough so that both $C_{L(xy)}$ and $C_{L(z)}$ contributed significantly to C_L below 2°K, then $C_{L(xy)}$ would still be appreciable, and proportional to T^3, between 2 and 4°K. Since we observe no such cubic term in this region, we assume that below 4°K $C_L\approx C_{L(z)}$ and find from Eqs. (14) and (11)

$$n_z=28.5\times10^{-6},\qquad \theta_{1z}=7.8°\text{K};\qquad \theta_{mz}=1020°\text{K}.$$

This value of θ_{mz} is somewhat higher than that estimated by Krumhansl and Brooks (950°K). Their

[30] R. O. Brennan, J. Chem. Phys. **20**, 40 (1952).
[31] K. Lonsdale, Proc. Phys. Soc. (London) **54**, 314 (1942).

[32] It has been pointed out by Newell (unpublished) that this assumption may be grossly in error, since it depends not merely on the geometry of the lattice but also on the interatomic forces connected with the two types of mode.

estimates of θ_{1z} and n_z (60°K, 5×10^{-2}, respectively) are much higher than the values calculated from the experimental data.

Although the experimental error in the measured points is only about 1–2%, a term proportional to T^3 between 2 and 4°K which was as large as 10% of C_L at 4°K would not be inconsistent with the line we have drawn in this temperature region. Taking this as an approximate upper limit for $C_{L(xy)}$ below 4°K, we find, using the relation due to Krumhansl and Brooks,

$$2n_{xy}\approx(\nu_{1xy}/\nu_{mxy})^2=(\theta_{1xy}/\theta_{mxy})^2, \tag{16}$$

that $\theta_{1xy}\approx20$°K, $n_{xy}\approx20\times10^{-6}$.

It should be mentioned that it is also possible to fit the data with the assumption that T_{0xy} is also about 2°K. On this assumption, θ_{1xy} is about equal to θ_{1z} at ~7–8°K, since the latter is not very sensitive to the assumption made concerning $C_{L(xy)}$. The major change is in the n's: n_z becomes 19×10^{-6} and n_{xy} becomes 10×10^{-6}. In either case, the data between 10 and 20°K correspond to a decrease in θ_{mz} in this temperature interval, if it is assumed that θ_{mxy} is roughly constant there. As has been pointed out by Krumhansl and Brooks, this decrease in θ_{mz} is to be expected as a consequence of the existence of peaks in the vibration spectrum[7,9] which are neglected in their treatment.

In order to account for the discrepancies between the predictions of the Krumhansl-Brooks analysis and the values of their parameters calculated from the experimental data, several possibilities suggest themselves:

(1) The decomposition into z modes and xy modes may not be appropriate; one of the other decompositions which have been suggested (Komatsu and Nagamiya,[3] Klemens,[6] Gurney[5]) might be more suitable.

(2) This decomposition may be suitable, but there may be difficulties in its application: (a) The values chosen for the force constants may be incorrect. These determine ν_{1z} and ν_{1xy} and hence the transition temperatures, T_{0z} and T_{0xy}. (b) The approximation of the spectrum below ν_{1z} by the Debye parabolic form may be too crude. The region of the Brillouin zone just below ν_{1z}, where the constant-energy surfaces are not yet truncated but have their upper portions crowded together near the upper planes of the zone, would not be likely to give rise to a parabolic spectrum. The transition should depend strongly, however, on this region of the zone. (c) There may be a size effect, due to the limited number of planes (60–90) in each crystallite.[23] For waves propagating in a direction making the angle θ with the z axis, the distance in the z direction between points of equal phase is given by $\lambda_z=\lambda/\cos\theta$. If λ_z is less than the crystallite thickness, L_z, the waves may not be transmitted coherently across the crystallite boundaries.[24] Wave vectors corresponding to $\nu>\nu_{1z}$ occupy a region of the Brillouin zone for which θ cannot have arbitrarily small values and so are not affected by the above argument. Their vibration spectrum will still be linear and their contribution to C_L should still be given by Eq. (14), above T_0 as calculated by Krumhansl and Brooks. But among the waves with $\nu<\nu_{1z}$ only those with sufficiently long wavelength could be transmitted through the material without attenuation.[25] The condition for such transmission would be $\lambda_z>L_z$ and these waves would be expected to have a vibration spectrum proportional to ν^2 and thus give a T^3 contribution to C_L, since for them the material would effectively be a continuum. There would, however, be a new transition temperature $T_0'<T_0$.

A very rough quantitative check of this idea can be made by using the relation

$$(h\bar{v}/\lambda_{zm})=k\theta_D, \tag{17}$$

between $\bar{v}$, the mean velocity of sound and λ_{zm}, the minimum wavelength for such "continuum waves." If we associate λ_{zm} with $L_z\approx250$ A and take $T_0'=2$°K, relating it to θ_D by $\theta_D=5T_0'$ (since C_L is proportional to T^3 below T_0'), then $\bar{v}$ is about 5×10^5 cm/sec. This is of the same order of magnitude as for an isotropic medium with the compressibility calculated for graphite by Brennan and with Poisson's ratio about 0.3.

Of these possibilities, (1) and (2b) involve further examination of the theory, while (2a) and (2c) would seem to susceptible to experimental investigation. In order to investigate thoroughly the possibility of a size effect it would be desirable to measure the heat capacity of a single crystal. Pure single crystals large enough for low-temperature heat capacity measurements unfortunately do not seem to be available at present.

We should like to express our appreciation of the valuable support and encouragement of Dr. K. Lark-Horovitz.

Note added in proof.—K. Komatsu [J. Phys. Soc. Japan **10**, 346 (1955)] has recently published a new analysis of the lattice atomic heat in which he predicts a transition from T^3 to λ^3 dependence at about 2°K.

[23] Gurney (reference 5) also discusses a size effect in connection with his model.

[24] A similar effect arising in thermal conduction in glass has been discussed by C. Kittel, Phys. Rev. **75**, 972 (1949).

[25] The same situation can also be discussed in terms of the modified Brillouin zone appropriate to crystallites of limited thickness. The general result has been described by P. P. Ewald, Proc. Phys. Soc. (London) **52**, 167 (1940).

Density Effect in Ionization Energy Loss of Charged Particles

JACOB NEUFELD AND R. H. RITCHIE
Oak Ridge National Laboratory, Oak Ridge, Tennessee

A more rigorous justification is given for the statement made by Aage Bohr that for an incident particle of relativistic velocity, the electrostatic effect of the surrounding electrons is negligible and the main factor in the screening is due to the electromagnetic interactions which limit the impact parameter to a value c/Ω, where $\Omega=(4\pi ne^2/m)^{\frac{1}{2}}$ and n is the electron density in the medium.

I. INTRODUCTION

THE density effect, originally suggested by Swann,[1] accounts for the reduction in the ionization energy loss of charged particles due to the polarization of the medium, and a quantitative theory of this effect has been given by Fermi.[2] The work of Fermi has been further extended and refined and a great many investigations have been devoted to this subject.[3] The density effect has also been studied by Bohr,[4] who emphasized certain intuitional and microscopic aspects whereas previous investigations were based mostly on rigorous solution of Maxwell's equations in a macroscopic medium. Bohr was particularly interested in the adiabatic limit of the impact parameter beyond which no appreciable energy transfer occurs, and showed by means of ingenious microscopic considerations how the electromagnetic field of the perturbed electrons is effective in setting this adiabatic limit.

The arguments presented by Bohr are based to a large extent on rather cursory calculations. It seemed, therefore, desirable to undertake more exact calculations to show whether and to what extent the Bohr theory is valid. The results obtained in this investigation show a satisfactory agreement with the results obtained by Bohr.

II. ARGUMENTS OF BOHR

Bohr's considerations are based on the existence of a maximum impact parameter defining the adiabatic limit beyond which the effect of the moving particle is negligible. For relativistic velocities of the particle, i.e., for $V\rightarrow c$, for a medium of low density, the adiabatic limit d_1 increases indefinitely with V, i.e.,

$$d_1=V\gamma/\omega_1, \tag{1}$$

where $\gamma=1/(1-\beta^2)^{\frac{1}{2}}$. We assume that the medium is composed of atoms characterized by a binding frequency ω_1.

In case of a dense medium the situation changes entirely. According to Bohr, electromagnetic forces enter into the effect caused by the electrons in the medium that are perturbed by the moving particle. As shown by Bohr, these forces limit the impact parameter, which for increasing V tends to a value

$$d_2=c/\Omega, \tag{2}$$

where $\Omega=(4\pi ne^2/m)^{\frac{1}{2}}$ and n is the electron density in the medium.

The expression (2) has been derived by considering an electron at a distance ρ from the particle track. Bohr[5] calculated that the field due to the retarded action of all electrons in the medium perturbed by the particles and acting on the electron at a distance ρ is:

$$E_{\text{ret}}{}^{m}\sim(e^2n\rho^2/c^3)\ddot{\eta}, \tag{3}$$

where $\ddot{\eta}$ is the acceleration of the electron due to the field of the particle. The field

$$E^{p}=m\ddot{\eta}/e, \tag{4}$$

and is directed against $E_{\text{ret}}{}^{m}$. Therefore, as observed by Bohr, we obtain from (3) and (4) that

$$E_{\text{ret}}{}^{m}\ll E^{p}, \tag{5}$$

for $\rho\ll c/\Omega$.

From this observation Bohr concluded that $\rho=c/\Omega$ represents the adiabatic limit beyond which the interaction of the incident particle with the surrounding medium is negligible.[5]

The above argument presented by Bohr is rather cursory. No consideration has been given to the impact parameter $\rho>c/\Omega$. In order to support Bohr's result, it would be necessary to show that for $\rho>c/\Omega$ the field produced by the retarded action of the perturbed electrons is equal and opposite to the field due to the direct action of the incident particle and, therefore, the medium remains unperturbed. This has not been done by Bohr.

It should be noted that the expression (3) is not rigorous and cannot be applied for $\rho>c/\Omega$. This can be shown by taking[6]

$$\ddot{\eta}\sim Z_1e^2\gamma/\rho^2, \tag{6}$$

(where Z_1e is the charge of the incident particle) and substituting the expression (6) in (3) we obtain that for any substantially large impact parameter ρ the field $E_{\text{ret}}{}^{m}$ is constant, i.e.,

$$E_{\text{ret}}{}^{m}\sim Z_1e^2n\gamma/c^2, \tag{7}$$

[1] W. F. G. Swann, J. Franklin Inst. **226**, 598 (1938).
[2] E. Fermi, Phys. Rev. **57**, 485 (1940).
[3] For a review of the literature, see, for instance, F. Sauter, in *Kosmische Strahlung*, edited by W. Heisenberg (Springer Verlag, Berlin, 1953), pp. 456–481.
[4] Aage Bohr, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **24**, No. 19 (1949).
[5] See reference 4, p. 23.
[6] See reference 4, p. 8, Formula (3.1).

which obviously is not in agreement with the physical reality.

III. EXACT CALCULATIONS

As pointed out by Bohr, the screening effect of the medium on the incident charged particle depends on the velocity of the particle. For a low velocity, one has to consider only the electrostatic forces between the electrons of the medium and these forces are only of secondary importance in the stopping power. For a relativistic velocity, the electrostatic effects are negligible and the stopping power is determined mainly by electromagnetic interactions between the electrons in the medium.

In order to determine separately the electrostatic and electromagnetic effects, we shall examine the longitudinal and transverse components of the field. Since the longitudinal component is irrotational, it will be associated with electrostatic effects and since the transverse component is divergence-free, it will be associated with electromagnetic interactions. It should be noted in that connection that the retarded field of the perturbed electrons which has been calculated by Bohr is not identical to the transverse component of the electromagnetic field.

To calculate in a precise manner the effect of the transverse and longitudinal fields in the dielectric medium we begin with Maxwell's equations:

$$\begin{aligned} &-\nabla\times\mathbf{E}=\frac{1}{c}\frac{\partial}{\partial t}\mathbf{H};\quad \nabla\cdot\mathbf{H}=0;\\ &\nabla\times\mathbf{H}=\frac{\epsilon}{c}\frac{\partial\mathbf{E}}{\partial t}+\frac{4\pi}{c}Z_1e\mathbf{V}\delta(\mathbf{r}-\mathbf{V}t), \qquad (8)\\ &\epsilon\nabla\cdot\mathbf{E}=4\pi Z_1e\delta(\mathbf{r}-\mathbf{V}t), \end{aligned}$$

in which we assume the medium to be infinite homogeneous, nonmagnetic, and characterized by a dielectric constant $\epsilon(\omega)$.

We now analyze $\mathbf{E}$, $\mathbf{H}$, and the incident charge into Fourier components,

$$\begin{Bmatrix}\mathbf{E}(\mathbf{r},t)\\ \mathbf{H}(\mathbf{r},t)\\ \delta(\mathbf{r}-\mathbf{V}t)\end{Bmatrix}=\int d\mathbf{k}\exp[i\mathbf{k}(\mathbf{r}-\mathbf{V}t)]\begin{Bmatrix}\mathbf{E}\\ \mathbf{H}\\ (2\pi)^{-3}\end{Bmatrix}, \qquad (9)$$

so that Maxwell's equations become

$$\begin{aligned} &i\mathbf{k}\times\mathbf{E}=i\omega\mathbf{H}/c,\quad \mathbf{k}\cdot\mathbf{H}=0,\\ &i\mathbf{k}\times\mathbf{H}=-i\omega\epsilon(\omega)\mathbf{E}/c+Z_1e\mathbf{V}/2\pi^2c, \qquad (10)\\ &\epsilon i\mathbf{k}\cdot\mathbf{E}=Z_1e/2\pi^2. \end{aligned}$$

Solving for $\mathbf{E}$ and dividing it into two components, $\mathbf{E}_l$ and $\mathbf{E}_t$, parallel and perpendicular, respectively, to the propagation vector we find

$$\mathbf{E}_l=-\frac{iZ_1e\mathbf{k}}{2\pi^2k^2\epsilon(\omega)}, \qquad (11)$$

$$\mathbf{E}_t=\frac{iZ_1e\omega}{2\pi^2c^2}\frac{\mathbf{V}-(\mathbf{k}\cdot\mathbf{V})\mathbf{k}/k^2}{(k^2-\omega^2\epsilon(\omega)/c^2)}. \qquad (12)$$

Going back into configuration space, we first divide the vector $\mathbf{k}$ into components ω/V and κ, parallel and perpendicular, respectively, to the particle track,

$$k^2=\kappa^2+\omega^2/V^2;\quad d\mathbf{k}=\kappa d\kappa d\varphi' d\omega/V. \qquad (13)$$

We find

$$\mathbf{E}_l(z,\rho,\varphi)=\frac{-iZ_1e}{2\pi^2V}\int_0^{2\pi}d\varphi'\int_0^{\infty}\kappa d\kappa\int_{-\infty}^{\infty}\frac{d\omega\mathbf{k}\exp[ik\rho\cos(\varphi-\varphi')+i\omega(z-Vt)/V]}{(\kappa^2+\omega^2/V^2)\epsilon(\omega)}, \qquad (14)$$

and

$$\mathbf{E}_t(z,\rho,\varphi)=\frac{iZ_1e}{2\pi^2c^2V}\int_0^{2\pi}d\varphi'\int_0^{\infty}\kappa d\kappa\int_{-\infty}^{\infty}\frac{\omega d\omega\{\exp[ik\rho\cos(\varphi-\varphi')+i\omega(z-Vt)/V]\}[\mathbf{V}k^2-\omega\mathbf{k}]}{(\kappa^2+\omega^2/V^2)\{\kappa^2+\omega^2[1-\beta^2\epsilon(\omega)]/V^2\}}, \qquad (15)$$

where (z,ρ,φ) are the coordinates of the field point in a cylindrical system of coordinates whose axis is coincident with the particle track. The location of the incident particle at time t is given by $z-Vt=0$.

An electron in the medium located at a distance ρ from the particle track is in the process of undergoing collision with the particle and simultaneously is acted upon by other electrons in the medium. We determine the net fields acting on an electron at ρ and at $z=Vt$ and, then, by carrying out the integrals given in Eqs. (14) and (15), we find that only the components of the longitudinal and transverse fields which are perpendicular to the track are not zero at the field point in question. We designate these by $E_l(\rho)$ and $E_t(\rho)$, respectively. Also let

$$\mathbf{E}(\rho)=\mathbf{E}_l(\rho)+\mathbf{E}_t(\rho). \qquad (16)$$

The symbols E_t and E_l shall be provided with superscripts p and m designating field generated by the "particle" and the "medium", respectively. By the straightforward application of Maxwell's equations to our dispersive medium, we obtain the following results that are valid for relativistic velocities and for a dielectric constant obtained from harmonic oscillator model.

The electric field due to the particle alone is

$$E_t^p(\rho)=Z_1e(1-\gamma)\frac{d}{d\rho}\left(\frac{1}{\rho}\right), \tag{17}$$

$$E_l^p(\rho)=-Z_1e\frac{d}{d\rho}\left(\frac{1}{\rho}\right), \tag{18}$$

$$E^p(\rho)=-\gamma Z_1e\frac{d}{d\rho}\left(\frac{1}{\rho}\right). \tag{19}$$

The electric field due to the action of the medium is

$$E_t^m(\rho)=\frac{d}{d\rho}\left[\Phi(\rho)-\frac{Z_1e}{\rho}\right]+\gamma Z_1e\frac{d}{d\rho}\frac{1-\exp(-\Omega\rho/V)}{\rho}, \tag{20}$$

$$E_l^m(\rho)=\frac{d}{d\rho}\left[\frac{Z_1e}{\rho}-\Phi(\rho)\right], \tag{21}$$

$$E^m(\rho)=\gamma Z_1e\frac{d}{d\rho}\frac{1-\exp(-\Omega\rho/V)}{\rho}, \tag{22}$$

where $\Phi(\rho)$ is the scalar potential in the medium. $\Phi(\rho)$ represents the Coulomb potential of the particle for points close to the particle. The value of $\Phi(\rho)$ for points distant from the particle is given in the Appendix, Eqs. (A-3) to (A-6).

We obtain from (22) that for small values of ρ the electric field exerted by the medium is

$$E^m\sim\gamma Z_1e\left(-\frac{\Omega^2}{2V^2}+\frac{\Omega^3\rho}{3V^3}-\frac{\Omega^4\rho^2}{8V^4}+\cdots\right). \tag{23}$$

We shall retain the first term in this expansion and compare it with the electric field due to the particle alone as given by (19). The field due to the medium acts in the direction opposite to the field of the particle. We obtain, however, that for $\rho\ll V/\Omega$,

$$E^m\ll E^p. \tag{24}$$

Consequently, for small impact parameters the shielding effect of the medium is negligible and the inequality (24) derived on rigorous grounds is in agreement with Bohr's inequality.

Consider now the region of large impact parameters corresponding to $\rho\gg V/\Omega$. Referring to (19) and (22), it is seen that for this region we have

$$E^m=-E^p, \tag{25}$$

and, consequently, the shielding effect is complete—i.e., it cancels entirely any effect due to the charged incident particle. The inequality (25) expresses the effect of the total electric field. However, it is apparent from (20) and (22) that for $V\to c$ (or for γ increasing indefinitely), $E_t^m\to E^m$ and, consequently, we may conclude that the shielding is caused primarily by the electromagnetic interaction of the electrons in the medium.

The authors wish to express their gratitude to Dr. Aage Bohr for an interesting conversation on the above subject and comments.

APPENDIX

For the sake of completeness, we shall give the expressions for longitudinal and transverse components of the electric field at all positions relative to the incident relativistic charged particle. The assumptions are, as stated before, (1) that the dielectric constant may be adequately represented by the usual expression

$$\epsilon(\omega)=1+\frac{4\pi Ne^2}{m}\sum_i\frac{f_i}{\omega_i^2-ig_i\omega-\omega^2},$$

where N is the number of atoms per cm^3 in the medium, f_i, ω_i, and g_i are the oscillator strength, resonant frequency, and damping constant belonging to the ith atomic transition, and (2) that the particle is in the relativistic region, i.e., that $\gamma=1/(1-\beta^2)^{\frac{1}{2}}\gg1$. The fields given here, in fact, comprise the first two terms of expansions in inverse powers of γ, e.g.,

$$E_t=\gamma E_t^{(1)}+E_t^{(2)}+O(\gamma^{-1}).$$

Using the formulation given in Eqs. (14) and (15), one finds for the components of the fields in the ρ direction (perpendicular to the particle track):

$$E_{t\rho}=\frac{\partial}{\partial\rho}\left\{\Phi(r)-\frac{\gamma Z_1e\exp[\Omega(\rho^2+\gamma^2z^2)^{\frac{1}{2}}/V]}{(\rho^2+\gamma^2z^2)^{\frac{1}{2}}}\right\},$$

$$+(E_{t\rho})_{\text{Čerenkov}} \tag{A-1}$$

$$E_{l\rho}=-\frac{\partial}{\partial\rho}\Phi(r), \tag{A-2}$$

where the scalar potential $\Phi(r)$ has the following forms:

(a) For field points close to the incident particle,

$$\Phi(r)=z_1e/r, \quad \text{where} \quad r=(z^2+\rho^2)^{\frac{1}{2}}. \tag{A-3}$$

(b) If the field point is far from the particle and if none of the ω_i are zero (the dielectric is an insulator),

$$\Phi(r)=\frac{1}{1+\Omega^2\sum_i(f_i/\omega_i^2)}\frac{Z_1e}{r}. \tag{A-4}$$

(c) If the field point is far from the particle and if one of the ω_i, say ω_1, is identically zero, then

$$\Phi(r)=-\frac{Z_1eg_1V}{f_1\Omega^2}\frac{d}{dz}\left(\frac{1}{r}\right). \tag{A-5}$$

(d) If the field point is far from the particle, $\omega_1=0$ and $g_1=0$

$$\Phi(r)=\frac{Z_1eV^2}{f_1\Omega^2}\frac{d^2}{dz^2}\left(\frac{1}{r}\right). \tag{A-6}$$

The z components (parallel with the particle track) are

$$E_{tz}=\frac{\partial}{\partial z}\left\{+\Phi(\mathbf{r})-\frac{Z_1e}{\gamma}\frac{\exp[-\Omega(\rho^2+\gamma^2z^2)^{\frac{1}{2}}/V]}{(\rho^2+\gamma^2z^2)}\right\}$$
$$-Z_1e\frac{\Omega^2}{V^2}\frac{z}{|z|}\int_{\gamma|z|}^{\infty}dx\frac{\exp[-\Omega(\rho^2+x^2)^{\frac{1}{2}}/V]}{(\rho^2+x^2)^{\frac{1}{2}}},$$
$$+(E_{tz})_{\text{Čerenkov}} \quad \text{(A-7)}$$

$$E_{tz}=-\frac{\partial}{\partial z}\Phi(\mathbf{r}). \quad \text{(A-8)}$$

The additional term in the expression for E_t corresponds to the existence of the Čerenkov wake behind the particle ($z<0$). Assuming for simplicity that only one of the resonances at $\omega=\omega_1$ is important in the expression for dielectric constant, we find that

$$(\mathbf{E}_t)_{\text{Čerenkov}}=\frac{2Z_1e\omega_1^2}{\Omega(\Omega^2+\omega_1^2)}\nabla(\omega_1z^2/\Omega^2-\rho^2)^{-\frac{1}{2}}$$
$$\text{if}\quad \omega_1z/\Omega<-\rho, \quad \text{(A-9)}$$

and

$$(\mathbf{E}_t)_{\text{Čerenkov}}=0 \quad \text{if}\quad \omega_1z/\Omega>-\rho. \quad \text{(A-10)}$$

These expressions are approximately valid as long as the field point is located sufficiently far from the surface of the cone $z=-\Omega\rho/\omega_1$. The apparent singularity on this cone does not exist but arises from the approximation to the more exact expression for $(\mathbf{E}_t)_{\text{Čerenkov}}$. It is to be noted that the angle of opening of the cone in the expression for E_t is

$$\vartheta=\tan^{-1}(\omega_1/\Omega). \quad \text{(A-11)}$$

The third term in (A-7) for E_{tz}, i.e.,

$$(E_{tz})_3=Z_1e\frac{\Omega^2}{V^2}\frac{z}{|z|}\int_{\gamma|z|}^{\infty}dx\frac{\exp[-\Omega(\rho^2+x^2)^{\frac{1}{2}}/V]}{(\rho^2+x^2)^{\frac{1}{2}}} \quad \text{(A-12)}$$

is of some interest and does not appear to have been noted before. This portion of the field is independent of the resonant frequency ω_1, but goes to zero for very low medium density. Exactly the same term arises in the transverse field expressions for a particle passing through a plasma.[7] This term is antisymmetric in the z coordinate and possesses a discontinuity at $z=0$. It arises from the Lorentz-contracted field of the line polarization charge which exists behind the incident particle.

[7] J. Neufeld and R. H. Ritchie, Phys. Rev. **98**, 1632 (1955).

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Dipolar Sums in the Primitive Cubic Lattices

M. H. Cohen, *Institute for the Study of Metals, University of Chicago, Chicago, Illinois, and Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania*

AND

F. Keffer, *Sarah Mellon Scaife Radiation Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania*

(Received March 11, 1955)

Dipole-wave sums, important in many magnetic and electric problems involving dipole-dipole interactions, are defined, and numerical values are given at sets of independent points in **k**-space equivalent to a 512-fold sampling of the first Brillouin zone of each of the three primitive cubic lattices. Strong size, shape, and position dependence of these sums is shown to occur in a pathological region about the origin in **k**-space. The dipole-wave sums are shown to be related to dipole-field factors at points within the unit cell. The dipolar anisotropy energy in the antiferromagnet MnO is discussed as an illustration of the use of dipole-wave sums.

I. INTRODUCTION

We introduce the lattice sums

$$S_n(\mathbf{k})=\rho^{-1}\sum_l{}'|\mathbf{r}_l|^{-n}\exp(i\mathbf{k}\cdot\mathbf{r}_l); \quad \text{(1a)}$$

$$S_n{}^{ij}(\mathbf{k})=\rho^{-1}\sum_l{}'r_l{}^ir_l{}^j|\mathbf{r}_l|^{-n}\exp(i\mathbf{k}\cdot\mathbf{r}_l). \quad \text{(1b)}$$

Here $i, j=x, y, z$; the primed sum is to be taken over all lattice vectors $\mathbf{r}_l$ except $\mathbf{r}_l=0$; and ρ is the number of lattice points per unit volume.

Specifically, we shall be interested in $S_3(\mathbf{k})$ and $S_3{}^{ij}(\mathbf{k})$, which we shall call *dipole-wave sums*. Our interest in these sums arises from their importance in many magnetic problems involving dipole-dipole interactions. In particular, we have made extensive use of dipole-wave sums in the quantum-mechanical problem of dipolar ferromagnetism.[1] Furthermore, these sums find use in the calculation of dipole-field factors (see Secs. VI, VII, and VIII) and of exciton energies.[2] They are important in the considerations of lattice stability of polar crystals that arise, for example, in connection with ferroelectrics and antiferroelectrics.

[1] M. H. Cohen and F. Keffer, Phys. Rev. **94**, 1412 (1954), and following paper [Phys. Rev. **99**, 1135 (1955)].

[2] W. R. Heller and A. Marcus, Phys. Rev. **84**, 809 (1951).

TABLE I. The choice of the unit of distance a for each lattice is made clear by the specification of the primitive lattice vectors in the table.

Direct lattice type	Direct lattice vectors $\mathbf{r}_1/a$	$\mathbf{r}_2/a$	$\mathbf{r}_3/a$	Reciprocal lattice vectors $\mathbf{b}_1a$	$\mathbf{b}_2a$	$\mathbf{b}_3a$	ρa^3	τa^2
Simple cubic (sc)	1,0,0	0,1,0	0,0,1	1,0,0	0,1,0	0,0,1	1	$\frac{1}{4}$
Face-centered cubic (fcc)	0,1,1	1,0,1	1,1,0	$\frac{1}{2},\frac{1}{2},\frac{1}{2}$	$\frac{1}{2},\frac{1}{2},\frac{1}{2}$	$\frac{1}{2},\frac{1}{2},\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
Body-centered cubic (bcc)	$\bar{1}$,1,1	1,$\bar{1}$,1	1,1,$\bar{1}$	0,$\frac{1}{2},\frac{1}{2}$	$\frac{1}{2}$,0,$\frac{1}{2}$	$\frac{1}{2},\frac{1}{2}$,0	$\frac{1}{2}$	$\frac{1}{4}$

II. METHOD OF OBTAINING DIPOLE-WAVE SUMS

A. Calculation of $S_n(\mathbf{k})$

Our notation conforms, in the main, to that introduced by Born and his students.[3-5] Some slight changes have been made, however, to suit our special purposes.

The most generally useful method of evaluating slowly convergent lattice sums is due to Ewald.[6] This method uses generalized theta functions and their transformation formula, and is most elegantly expounded by Born and Bradburn.[5] In particular, these authors show that for an infinite lattice [see their Eq. (4.7)]:

$$S_n(\mathbf{k})=\rho^{-1}[(\pi\tau)^{\frac{1}{2}n}/\Gamma(\tfrac{1}{2}n)][S_A+S_B-(2/n)], \quad (2)$$

with

$$S_A=\sum_l{}'[\exp(i\mathbf{k}\cdot\mathbf{r}_l)\varphi_{\frac{1}{2}n-1}(\tau r_l^2\pi)],$$
$$S_B=\rho\tau^{-1}\sum_l\varphi_{\frac{1}{2}-\frac{1}{2}n}[(2\pi\mathbf{b}_l-\mathbf{k})^2/4\pi\tau].$$

In this expression S_A is summed over the direct lattice, $\mathbf{r}_l$; S_B is summed over the reciprocal lattice, $\mathbf{b}_l$. The sum $S_n(\mathbf{k})$ is independent of the choice of Ewald's parameter τ, the idea being to choose this parameter for equally rapid convergence in the sums over the direct and the reciprocal lattices. The φ functions (introduced by Misra[3]) are defined by

$$\varphi_m(x)=\int_1^\infty \beta^m e^{-\beta^2 x}d\beta. \quad (3a)$$

These functions satisfy the recurrence formula

$$\varphi_m(x)=\varphi_0(x)+(m/x)\varphi_{m-1}(x); \quad (3b)$$

and, in particular,

$$\varphi_0(x)=e^{-x}/x,$$
$$\varphi_{-1}(x)=-\mathrm{Ei}(-x),$$
$$\varphi_{-\frac{1}{2}}(x)=(\pi/x)^{\frac{1}{2}}[1-\Phi(x^{\frac{1}{2}})].$$

Here $\mathrm{Ei}(-x)$ is the exponential integral, and $\Phi(x)$ is Gauss' error function.

Table I shows a set of lattice vectors and other parameters for the three primitive cubic lattices, expressed in terms of convenient units of distance.

B. Calculation of $S_3^{ij}(\mathbf{k})$

We note that

$$S_n^{ij}(\mathbf{k})=\frac{\partial}{\partial k^i}\frac{\partial}{\partial k^j}S_n(\mathbf{k}). \quad (4)$$

Using $S_3(\mathbf{k})$ as given by Eq. (2), we readily obtain $S_3^{ij}(\mathbf{k})$ from Eq. (4). In taking derivatives we observe that

$$\frac{\partial}{\partial x}\varphi_m(x)=-\varphi_{m+1}(x).$$

The result is

$$S_3^{ij}(\mathbf{k})=\rho^{-1}[(\pi\tau)^{5/2}/\Gamma(5/2)][S_C+S_D+\delta^{ij}S_E], \quad (5)$$

with

$$S_C=\sum_l{}'[r_l{}^ir_l{}^j\exp(i\mathbf{k}\cdot\mathbf{r}_l)\varphi_{\frac{1}{2}}(\tau r_l^2\pi)],$$
$$S_D=-(\rho/4\pi^2\tau^{7/2})\sum_l{}'[(2\pi b_l{}^i-k^i)(2\pi b_l{}^j-k^j)\times\varphi_0\{(2\pi\mathbf{b}_l-\mathbf{k})^2/4\pi\tau\}],$$
$$S_E=(\rho/2\pi\tau^{5/2})\sum_l\varphi_{-1}\{(2\pi\mathbf{b}_l-\mathbf{k})^2/4\pi\tau\}.$$

III. SELECTING A SAMPLE OF k-SPACE

The dipole-wave sums will be of use in numerical integrations (summations) over $\mathbf{k}$-space; hence the values of $\mathbf{k}$ for which sums are calculated must be chosen with care.

Since the sums are taken over all direct lattice points, their values will repeat in the various Brillouin zones (BZ) of $\mathbf{k}$-space. Thus, we need consider only the first zone. Furthermore, within this zone there will be considerable repetition of values due to cubic symmetry.

The choice of the $\mathbf{k}$ sampling is detailed below for the three cubic lattices.

A. Simple Cubic

If we define integers $q^i(i=x, y, z)$ by the relation

$$8k^i=(2\pi/a)q^i, \quad \text{(sc)} \quad (6)$$

we go from boundary to boundary of the first BZ on allowing q^i to vary from -4 to $+4$. Our sampling of $\mathbf{k}$-space thus consists of the 512 number triples (q^x,q^y,q^z), properly weighted at surfaces, edges, and corners of the BZ as discussed in Sec. IV. We need not evaluate 512 different sums, however, because of the following types of relationships imposed by cubic symmetry:

[3] R. D. Misra, Proc. Cambridge Phil. Soc. **36**, 173 (1940).
[4] M. Born and M. Bradburn, Proc. Cambridge Phil. Soc. **39**, 104 (1943).
[5] M. Bradburn, Proc. Cambridge Phil. Soc. **39**, 113 (1943).
[6] P. P. Ewald, Ann. Physik **64**, 253 (1921).

(1) Reflections, such as

$$S_3(l,m,n)=S_3(-l,m,n),$$
$$S_5^{xy}(l,m,n)=-S_5^{xy}(-l,m,n), \quad (7)$$
$$S_5^{xz}(l,m,n)=S_5^{xz}(-l,m,n).$$

(2) Permutations, such as

$$S_3(l,m,n)=S_3(m,l,n),$$
$$S_5^{xy}(l,m,n)=S_5^{yz}(m,l,n). \quad (8)$$

Thus we need evaluate sums for only the 35 points given by

$$4\geq q^x\geq q^y\geq q^z\geq 0. \quad \text{(sc)} \quad (9)$$

B. Face-Centered Cubic

Here we define integers q^i by the relation

$$16k^i=(2\pi/a)q^i, \quad \text{(fcc)} \quad (10)$$

with q^i varying from -8 to $+8$.

Again the symmetry relationships of Eqs. (7) and (8) reduce the number of points we need consider.

The first BZ is now to be pictured in the reciprocal bcc lattice. If we stretch this reciprocal lattice by the scale factor 2π, the first BZ is then a Wigner-Seitz unit cell of a bcc. In order to stay inside this zone we must further restrict ourselves to points such that $q^x+q^y+q^z\leq 12$. This may also be seen from consideration of inversion symmetry and the fact that in our bcc the point $(8+l, 8+m, 8+n)$ is equivalent to (l,m,n).

Furthermore, since the reciprocal lattice is a bcc we must sample it with a bcc-like array of points; this restricts us to q^x, q^y, q^z all odd or all even.

Thus we need evaluate sums for only the 30 points given by

$$8\geq q^x\geq q^y\geq q^z\geq 0,$$
$$q^x+q^y+q^z\leq 12, \quad \text{(fcc)} \quad (11)$$
$$q^x,\ q^y,\ q^z \text{ all odd or all even.}$$

The equivalence of the boundary points (8,2,2) and (6,6,0) further reduces us to 29 points.

C. Body-Centered Cubic

This is somewhat similar to the fcc. We define integers q^i by

$$16k^i=(2\pi/a)q^i. \quad \text{(bcc)} \quad (12)$$

The first BZ is now a Wigner-Seitz unit cell of an fcc lattice. We fill this with an fcc-like array of points, and need evaluate sums for only the 29 points given by

$$8\geq q^x\geq q^y\geq q^z\geq 0,$$
$$q^x+q^y\leq 8, \quad \text{(bcc)} \quad (13)$$
$$q^x+q^y+q^z \text{ even.}$$

IV. NUMERICAL INTEGRATION IN k-SPACE

In applications of our work, sums like

$$\sum_{\mathbf{k}'} f(\mathbf{k}')=N\bar{f} \quad (14)$$

occur.[1] The summation is carried out over the N values of $\mathbf{k}'$ inside the first BZ which satisfy appropriate boundary conditions. Since $f(\mathbf{k}')$ is known only at certain points $\mathbf{k}$, the actual average, $\bar{f}$, must be replaced by an approximate average, $\langle f\rangle$, obtained by numerical integration.

The $\mathbf{k}$'s chosen in III reproduce the reciprocal lattice (times 2π) on a scale 8 times smaller. Each $\mathbf{k}$ may, therefore, be thought to lie at the center of a cell geometrically similar to, but 512 times smaller than, the first BZ. We then replace the average value of f in a small cell by the value of f at the center in performing the numerical integration. Cells at faces, edges, and corners of the first BZ do not lie entirely within it and must be assigned fractional weights, $W_{\mathbf{k}}$. Thus,

$$\langle f\rangle=(512)^{-1}\sum_{\mathbf{k}} f(\mathbf{k})W_{\mathbf{k}}. \quad (15)$$

The summation in (15) should, of course, be performed first over those $\mathbf{k}$'s for which $f(\mathbf{k})$ is the same, in particular for $\mathbf{k}$'s connected by 2π times a reciprocal lattice vector.

A list of face, edge, and corner points and their weights is given in Table II. All points related by a permutation or reflection have the same weight.

V. BEHAVIOR NEAR k=0

It turns out that the sums $S_3(\mathbf{k})$ and $S_5^{ij}(\mathbf{k})$ for both finite and infinite lattices are rapidly varying functions of $\mathbf{k}$ in the vicinity of the point $\mathbf{k}=0$. We study the small $\mathbf{k}$ behavior in this section because of its importance in specific applications, e.g., in ferromagnetic resonance, and in obtaining representative values for numerical integrations.

A. Infinite Lattice

To avoid divergences we form the combination

$$D^{ij}(\mathbf{k})=3S_5^{ij}(\mathbf{k})-S_3(\mathbf{k})\delta^{ij}. \quad (16)$$

We note that $D^{ij}(0)$ is the conventional dipole-field factor for lattice points. Direct evaluation yields

$$D^{ij}(0)=0. \quad (17)$$

On the other hand, an expression for $D^{ij}(\mathbf{k})$ valid in the limit of small $\mathbf{k}$ may be obtained by manipulation of Eqs. (2) and (5),

$$D^{ij}(\mathbf{k})=4\pi^2\rho^{-1}\tau^{5/2}\{S_C+S_D+(2\pi\tau)^{-1}(\tfrac{2}{3}-S_A)\delta^{ij}\};$$

$$\lim_{\mathbf{k}\to 0} S_A(\mathbf{k})=\sum_l{}' \varphi_{\frac{1}{2}}(\pi r_l^2\tau);$$

$$\lim_{\mathbf{k}\to 0} S_C(\mathbf{k})=\tfrac{1}{3}\delta^{ij}\sum_l{}' r_l^2\varphi_{\frac{1}{2}}(\pi r_l^2\tau);$$

$$\lim_{\mathbf{k}\to 0} S_D(\mathbf{k})=-\tfrac{1}{3}\rho\delta^{ij}\tau^{-7/2}\sum_l{}' b_l^2\varphi_0(\pi b_l^2\tau)$$
$$-(\rho/4\pi^2\tau^{7/2})k^ik^j\varphi_0(k^2/4\pi\tau).$$

Putting these together, using the recurrence formula (3b) for the φ's, and using Poisson's summation for-

mula,[7] one obtains

$$\lim_{\mathbf{k}\to 0} D^{ij}(\mathbf{k}) = \rho(4\pi/3)[\delta^{ij} - 3(k^i k^j/k^2)]. \quad (18)$$

Equation (18) is structure-independent and in fact could have been obtained by replacing the sum by an integral and surrounding the origin by a sphere of exclusion.[8] Comparing Eq. (17) and Eq. (18) we see that $D^{ij}(\mathbf{k})$ is only piecewise continuous at $\mathbf{k}=0$ and that the average of $D^{ij}(\mathbf{k})$ in the neighborhood of $\mathbf{k}=0$ is equal to $D^{ij}(0)$.

Now the original definition of $D^{ij}(\mathbf{k})$,

$$D^{ij}(\mathbf{k}) = \rho^{-1} \sum_l{}' (3r_l{}^i r_l{}^j - r_l{}^2\delta^{ij}) r_l{}^{-5} \exp(i\mathbf{k}\cdot\mathbf{r}_l), \quad (19)$$

is in fact the explicit Fourier series for $D^{ij}(\mathbf{k})$ which is periodic in $\mathbf{k}$. We might therefore expect that a study of the convergence of Eq. (19) near $\mathbf{k}=0$ would throw some light on its piecewise continuity. Accordingly we turn to a discussion of $D^{ij}(\mathbf{k})$ in finite lattices.

B. Pathological Behavior Near $\mathbf{k}=0$ in Finite Lattices

In finite lattices the $D^{ij}(\mathbf{k})$ are not as directly useful at very small $\mathbf{k}$ as are other dipolar sums more directly related to the geometry of the sample. For example, for a long thin cylinder the plane-wave factor in the sum can be replaced by a cylindrical wave. Nevertheless, certain difficulties in the behavior of dipole-wave sums near small $\mathbf{k}$ occur for all types of waves. To illustrate these difficulties we now calculate $D^{ij}(\mathbf{k})$ for a finite lattice of given shape, chosen spherical for convenience. The finite sum is related to the infinite sum by

$$\sum_V = \sum_\infty - \sum_{\infty - V}. \quad (20)$$

The sum $\sum_{\infty-V}$ may be converted into an integral for small $\mathbf{k}$ and large volume V. Hence for a sphere of radius R, taking the origin of the sum in the center of the sphere,

$$D^{ij}(\mathbf{k}) = \frac{4\pi}{3}\left(\delta^{ij} - \frac{3k^i k^j}{k^2}\right) - \int_R^\infty \frac{3r^i r^j - r^2}{r^5} \exp(i\mathbf{k}\cdot\mathbf{r}) d^3r. \quad (21)$$

The integral is easily evaluated and leads to

$$D^{ij}(\mathbf{k}) = \frac{4\pi}{3}\left(\delta^{ij} - \frac{3k^i k^j}{k^2}\right)\left[1 - \frac{3j_1(kR)}{kR}\right], \quad (22)$$

which goes to zero as $k \to 0$. Thus, as it must, the discontinuity disappears for the finite lattice. It becomes instead a rapid variation of $D^{ij}(\mathbf{k})$ near $\mathbf{k}=0$, a fact which must be recognized in physical applications of the $D^{ij}(\mathbf{k})$.

TABLE II. Weights of face, edge, and corner points.

sc		fcc		bcc	
q	W_q	q	W_q	q	W_q
$4q_xq_z$	1	444	1	440	1
$4q_z$	1	642	1	442	1
444	1	800	1	530	1
		820	1	532	1
		660	1	620	1
		822	1	710	1
		840	1	622	1
				444	1
				800	1

One sees from Eq. (22) that the finite sum differs from the infinite sum only for $\mathbf{k}$'s such that $kR \lesssim 10$. Since the separation $\Delta\mathbf{k}$ between $\mathbf{k}$'s in the numerical integration is such that $\Delta k a \sim 1$, where a = lattice constant, this difference between the finite and the infinite sums is appreciable only over an extremely small region of the cell about $\mathbf{k}=0$. We shall call this region the *pathological region*.

Further, the value for the finite lattice, $D^{ij}(0)=0$ for a sphere, represents very poorly the region about $\mathbf{k}=0$ in the numerical integration, a consequence of the rapid variation of $D^{ij}(\mathbf{k})$ near $\mathbf{k}=0$. A more representative value would be obtained by using Eq. (18) alone or together with the next terms in an expansion in powers of $\mathbf{k}$ of $D^{ij}(\mathbf{k})$ for an *infinite* lattice and averaging over a sphere approximating the $\mathbf{k}=0$ cell. An accurate value is not important since there are 512 cells in the numerical integration, with inaccuracy only at $\mathbf{k}=0$.

The difference between the infinite and finite sums, $G^{ij}(\mathbf{k})$, is proportional to $j_1(kR)/kR$. Let ϵ be a small positive number. Then if Eq. (19) is uniformly convergent for all $\mathbf{k}$, it is possible to find a value of R such that for all radii greater than R, $|G^{ij}(\mathbf{k})| < \epsilon$ for all $\mathbf{k}$. But if we include the point $\mathbf{k}=0$ in the region of variability of $\mathbf{k}$, then for $k \ll 1/R$, $|G^{ij}(k)| \sim 1$ for radii $\gg R$. Thus the Fourier series (19) for $D^{ij}(\mathbf{k})$ in an infinite lattice is not uniformly convergent near $\mathbf{k}=0$. Hence $D^{ij}(\mathbf{k})$ cannot be a uniformly continuous function of $\mathbf{k}$. Since the Fourier series exists, the function must be piecewise continuous, a conclusion in accord with the results of the previous section.

Let us repeat for emphasis that the above considerations hold in detail only when the origin of the sum is at the center of the sphere. A similar treatment of the case when the origin of the sum lies at an arbitrary point within the sphere yields the result that the dipole-wave sums show a strong *position dependence* when $kR \lesssim 10$.

The dipole-wave sums rapidly become independent of origin as kR increases beyond 10. In the special case of $k=0$ the dipole-wave sums are completely independent of origin (except for origins in the layer immediately next the surface).

[7] In our particular case this is the theta-function transformation $r^4 \sum_l \exp(-\pi r_l{}^2 r) = \rho \sum_l \exp(-\pi b_l{}^2/r)$. See reference 5 or 6.

[8] See T. Holstein and H. Primakoff, Phys. Rev. 58, 1098 (1940).

TABLE III. Structure-dependent terms.

	fcc	bcc
α	4.004	4.124
β	1.237	0.849
γ^{zzzz}	−4.288	−3.487
γ^{zzyy}	−1.710	−1.591

Although all considerations have been made explicitly for a sphere, similar results have been obtained for an infinite cylinder and for an infinite flat plate. We expect, therefore, that such behavior occurs for all shapes.

To sum up, we find that $D^{ij}(\mathbf{k})$ is not a uniformly convergent series and has an associated discontinuity at $\mathbf{k}=0$ in the function it represents for an infinite lattice. A strong size, shape, and position dependence of the finite sum occurs for a very small region near $\mathbf{k}=0$. Finally, the *infinite* sum, appropriately weighted, may always be used in numerical integration over all $\mathbf{k}$-space, even for finite lattices.

C. Infinite Lattice, Structure Dependent Terms

In Sec. IV $\mathbf{k}$-space was divided into 512 cells for purposes of numerical integration, and the value of the dipole-wave sum at the center of each cell was used to approximate its average value over the cell. In Sec. V.B it was shown that finite-lattice behavior extends over a negligible region of the cell surrounding the origin, and it becomes necessary to consider the infinite lattice behavior in the central cell in more detail in order to find representative values for numerical integration. To do this we carry the expansion of the infinite sums further than in Sec. VA by including terms of order k^2. The terms of order k^2 to be added to Eq. (18) are

$$k^2a^2\left[\alpha\frac{k^ik^j}{k^2}+\beta\delta^{ij}+\gamma^{ijmn}\frac{k^mk^n}{k^2}\right], \tag{18'}$$

where

$$\alpha=\tau^{-1}\{1+\tfrac{1}{3}\textstyle\sum_l{}'\varphi_0(\pi b_l^2/\tau)[1+(4\pi b_l^2/\tau)]\};$$

$$\beta=(\pi\tau^{\frac{1}{2}}/3\rho)\textstyle\sum_l{}'r_l^2\varphi_{\frac{1}{2}}(\pi\tau r_l^2)+(1/3\tau)\sum_l{}'\varphi_0(\pi b_l^2/\tau)[1+(\pi b_l^2/\tau)];$$

$$\gamma^{ijmn}=-(2\pi^2\tau^{5/2}/\rho)\textstyle\sum_l{}'r_l{}^ir_l{}^jr_l{}^mr_l{}^n\varphi_{\frac{1}{2}}(\pi\tau r_l^2)-(4/\tau)\sum_l{}'\varphi_0(\pi b_l^2/\tau)b_l{}^ib_l{}^jb_l{}^mb_l{}^n\times[b_l^{-4}+(\pi^2/2\tau^2)+(\pi/\tau b_l^2)].$$

We have evaluated α, β, and the independent nonzero values of γ^{ijmn} for the fcc and bcc. These are given in Table III.

VI. TABULATIONS AND CHECKS

The calculated values of $S_2(\mathbf{k})$ and $S_4{}^{ij}(\mathbf{k})$ for our samplings of $\mathbf{k}$-space are given in Tables IV, V, and VI. The representative integers q^i are defined in Eqs. (6), (10), and (12).

TABLE IV. Simple cube.

$q^xq^yq^z$	S_2	$S_4{}^{xx}$	$S_4{}^{yy}$	$S_4{}^{zz}$	$S_4{}^{xy}$	$S_4{}^{yz}$	$S_4{}^{zx}$
100	12.465	1.298	5.584	5.584	0	0	0
110	8.391	2.067	2.067	4.257	−1.838	0	0
111	6.111	2.037	2.037	2.037	−1.151	−1.151	−1.151
200	4.678	−1.452	3.065	3.065	0	0	0
210	3.514	−0.962	1.771	2.706	−1.178	0	0
211	2.597	−0.671	1.634	1.634	−0.920	−0.920	−0.485
220	1.315	−0.363	−0.363	2.042	−1.153	0	0
221	0.772	−0.276	−0.276	1.325	−0.960	−0.514	−0.514
222	−0.405	−0.135	−0.135	−0.135	−0.602	−0.602	−0.602
300	1.247	−2.750	1.998	1.998	0	0	0
310	0.745	−2.428	1.317	1.857	−0.528	0	0
311	0.303	−2.170	1.236	1.236	−0.446	−0.446	−0.220
320	−0.363	−1.799	−0.114	1.551	−0.613	0	0
321	−0.688	−1.649	−0.093	1.055	−0.530	−0.289	−0.274
323	−1.434	−1.328	−0.053	−0.053	−0.361	−0.361	−0.351
330	−1.354	−1.318	−1.318	1.283	−0.363	0	0
331	−1.587	−1.237	−1.237	0.888	−0.320	−0.172	−0.172
332	−2.135	−1.055	−1.055	−0.025	−0.227	−0.226	−0.226
333	−2.664	−0.888	−0.888	−0.888	−0.149	−0.149	−0.149
400	0.222	−3.155	1.689	1.689	0	0	0
410	−0.134	−2.879	1.150	1.595	0	0	0
411	−0.461	−2.642	1.091	1.091	0	0	−0.153
420	−0.959	−2.282	−0.057	1.380	0	0	0
421	−1.218	−2.123	−0.048	0.954	0	0	−0.201
422	−1.827	−1.766	−0.030	−0.030	0	0	−0.268
430	−1.737	−1.767	−1.150	1.181	0	0	0
431	−1.937	−1.669	−1.091	0.823	0	0	−0.133
432	−2.413	−1.443	−0.953	−0.016	0	0	−0.179
433	−2.878	−1.232	−0.823	−0.823	0	0	−0.121
440	−2.046	−1.574	−1.574	1.102	0	0	0
441	−2.224	−1.498	−1.498	0.771	0	0	0
442	−2.649	−1.319	−1.319	−0.011	0	0	0
443	−3.067	−1.148	−1.148	−0.771	0	0	0
444	−3.239	−1.080	−1.080	−1.080	0	0	0

TABLE V. Face-centered cube.

$q^xq^yq^z$	S_k	S_k^{xx}	S_k^{yy}	S_k^{zz}	S_k^{xy}	S_k^{xz}	S_k^{yz}
111	11.093	3.698	3.698	3.698	−1.368	−1.368	−1.368
200	9.398	0.527	4.435	4.435	0	0	0
220	5.513	1.244	1.244	3.025	−1.958	0	0
222	3.447	1.149	1.149	1.149	−1.300	−1.300	−1.300
311	3.851	−0.384	2.118	2.118	−1.043	−1.043	−0.373
331	1.355	0.087	0.087	1.182	−1.662	−0.619	−0.619
333	0.157	0.052	0.052	0.052	−1.233	−1.233	−1.233
400	2.029	−1.463	1.746	1.746	0	0	0
420	1.101	−0.979	0.818	1.262	−1.357	0	0
422	0.462	−0.697	0.579	0.579	−1.147	−1.147	−0.712
440	−0.512	−0.394	−0.394	0.276	−1.422	0	0
442	−0.662	−0.330	−0.330	−0.003	−1.311	−0.876	−0.876
444	−0.810	−0.270	−0.270	−0.270	−1.205	−1.205	−1.205
511	−0.060	−1.595	0.767	0.767	−0.562	−0.562	−0.182
531	−0.755	−0.981	0.052	0.174	−1.161	−0.459	−0.414
533	−0.940	−0.592	−0.174	−0.174	−1.040	−1.040	−0.996
551	−1.432	−0.395	−0.395	−0.642	−0.980	−0.402	−0.402
600	−0.978	−1.979	0.501	0.501	0	0	0
620	−1.192	−1.605	0.226	0.188	−0.661	0	0
622	−1.284	−1.282	−0.001	−0.001	−0.619	−0.619	−0.540
640	−1.581	−0.846	−0.162	−0.572	−0.815	0	0
642	−1.375	−0.595	−0.186	−0.595	−0.804	−0.578	−0.804
660	−1.815	−0.227	−0.227	−1.359	−0.516	0	0
711	−1.661	−1.850	0.095	0.095	−0.181	−0.181	−0.146
731	−1.744	−1.224	−0.095	−0.425	−0.405	−0.167	−0.370
800	−1.824	−2.052	0.114	0.114	0	0	0
820	−1.871	−1.698	−0.0002	−0.172	0	0	0
840	−1.918	−0.902	−0.114	−0.902	0	0	0

We believe the accuracy of these tables to be within rounding-off error of the figures as given. Several interesting checks have been made, the most important of which is provided by the relation

$$S_k(\mathbf{k})=S_k^{xx}(\mathbf{k})+S_k^{yy}(\mathbf{k})+S_k^{zz}(\mathbf{k}). \tag{23}$$

The left and right sides of this relation were calculated independently and checked against each other only after all the numbers were tabulated. In this way the number of significant figures was established. To within rounding-off error of the numbers as given, no violations of Eq. (23) were found.

It is important to know how representative is our sampling of $\mathbf{k}$-space. To study this we have formed

TABLE VI. Body-centered cube.

$q^xq^yq^z$	S_k	S_k^{xx}	S_k^{yy}	S_k^{zz}	S_k^{xy}	S_k^{xz}	S_k^{yz}
110	10.760	2.925	2.925	4.190	−2.048	0	0
200	6.768	−0.231	3.500	3.500	0	0	0
211	4.598	0.356	2.121	2.121	−1.295	−1.295	−0.622
220	3.163	0.481	0.481	2.201	−1.935	0	0
222	1.350	0.451	0.451	0.451	−1.126	−1.126	−1.126
310	2.090	−1.045	1.390	1.745	−1.137	0	0
321	0.736	−0.605	0.325	1.016	−1.541	−0.697	−0.386
330	−0.078	−0.510	−0.510	0.941	−1.819	0	0
332	−0.674	−0.409	−0.409	0.145	−1.222	−0.594	−0.594
400	−0.111	−1.577	0.844	0.844	0	0	0
411	−0.230	−1.321	0.545	0.545	−0.755	−0.755	−0.076
420	−0.480	−1.141	−0.028	0.690	−1.411	0	0
422	−0.913	−0.883	−0.015	−0.015	−0.950	−0.950	−0.134
431	−0.969	−0.834	−0.545	0.412	−1.548	−0.286	−0.066
433	−1.439	−0.616	−0.411	−0.411	−0.612	−0.612	−0.060
440	−1.023	−0.787	−0.787	0.551	−1.770	0	0
442	−1.325	−0.659	−0.659	−0.006	−1.218	0	0
444	−1.619	−0.540	−0.540	−0.540	0	0	0
510	−1.344	−1.383	−0.073	0.112	−0.609	0	0
521	−1.424	−1.043	−0.418	0.038	−1.010	−0.408	+0.113
530	−1.276	−0.808	−0.808	0.341	−1.456	0	0
532	−1.461	−0.645	−0.645	−0.170	−0.995	−0.367	+0.367
600	−2.090	−1.221	−0.435	−0.434	0	0	0
611	−2.001	−1.029	−0.487	−0.487	−0.375	−0.375	+0.137
620	−1.848	−0.844	−0.844	−0.159	−0.782	0	0
622	−1.755	−0.585	−0.585	−0.585	−0.523	−0.523	+0.523
710	−2.369	−0.858	−0.858	−0.654	−0.210	0	0
800	−2.572	−0.857	−0.857	−0.857	0	0	0

TABLE VII. Lattice-point field factors of Luttinger and Tisza sc arrays.

Array	LT field factor	sc Dipole-wave sum	Value from Table IV
Z_1	0	$D^{zz}(0,0,0)$	0[a]
Z_2	−9.687	$D^{zz}(0,0,4)=D^{zz}(4,0,0)$	−9.687
Z_3	4.844	$D^{zz}(4,0,0)$	4.845[b]
Z_4	4.844	$D^{zz}(0,4,0)=D^{zz}(4,0,0)$	4.845
Z_5	5.351	$D^{zz}(4,4,0)$	5.352[c]
Z_6	−2.676	$D^{zz}(0,4,4)=D^{zz}(4,4,0)$	−2.676
Z_7	−2.676	$D^{zz}(4,0,4)=D^{zz}(4,4,0)$	−2.676
Z_8	0	$D^{zz}(4,4,4)$	0

[a] For a sphere.
[b] Since $D^{zz}(4,0,0)=D^{yy}(4,0,0)$ and $\sum_l D^{ij}(4,0,0)=0$, $D^{xx}(4,0,0)=-2D^{zz}(4,0,0)$. Therefore, the LT value is probably correct. Our error could have occurred in multiplying the rounding-off in S^{zz} by 3.
[c] Similarly, $D^{xx}(4,4,0)=-2D^{zz}(4,4,0)$. Therefore our value is probably correct.

sums over the N values of $\mathbf{k}'$ inside the first BZ of the type

$$N^{-1}\sum_{\mathbf{k}'}[S_{\mathbf{k}}{}^{ij}(\mathbf{k}')]^2=\rho^{-2}\sum_l{}'[r_l{}^i r_l{}^j|r_l|^{-5}]^2. \quad (24)$$

The right-hand side converges so rapidly it may be directly summed. The left-hand side, then, may be approximated from our tables. The agreement with the direct sums for six such approximate sums was found to hold to three significant figures—results of truly surprising accuracy. Further details on such sampling checks will be presented in a following paper.[1]

An interesting feature of the dipole-wave sums is that, for example, the sc value of $3S_{\mathbf{k}}{}^{zz}(0,0,4)-S_{\mathbf{k}}(0,0,4)=D^{zz}(0,0,4)$ is the z-component of the dipole-field factor at a lattice point of the Luttinger-Tisza[9] (LT) dipolar array Z_2. This array, which has alternate xy planes of dipoles pointing in the $+z$ and $-z$ directions, is one of the 8 basic arrays considered by LT. The lattice-point field factors for the other 7 arrays may also be obtained from appropriate dipole-wave sums. Table VII compares our values with those given by LT. Similar checks may be made for a few of the fcc and bcc arrays, but the results will not be given here. The agreement is excellent. In Sec. VII we show how it is possible to obtain from the dipole-wave sums the dipole-field factors at positions other than lattice points.

VII. DIPOLE-FIELD FACTORS FOR SITES WITHIN THE UNIT CELL

Consider the sum

$$D^{ij}(\mathbf{r}_m/\nu)=\sum_{\mathbf{k}}D^{ij}(\mathbf{k})W_{\mathbf{k}}\exp(i\mathbf{k}\cdot\mathbf{r}_m). \quad (25)$$

One can show by inserting Eq. (19) into Eq. (25) and inverting the order of summation that $D^{ij}(\mathbf{r}_m/\nu)$ is the dipole-field factor at the point $\mathbf{r}_m/\nu$ within the unit cell. That is,

$$\rho D^{ij}(\mathbf{r}_m/\nu)=\sum_l{}'\frac{3(r_l{}^i-r_m{}^i\nu^{-1})(r_l{}^j-r_m{}^j\nu^{-1})-(\mathbf{r}_l-\mathbf{r}_m\nu^{-1})^2\delta^{ij}}{|\mathbf{r}_l-\mathbf{r}_m\nu^{-1}|^5}. \quad (26)$$

Here $\mathbf{r}_m$ is an appropriately selected lattice vector, and ν is the ratio of the separation between two points of the reciprocal lattice and the corresponding points in the sampling of $\mathbf{k}$-space. For our choice of points, $\nu=8$. Eq. (25) may be inverted to express dipole-wave sums in terms of dipole-field factors.

Thus the values of dipole-wave sums tabulated in this paper can be used to obtain dipole-field factors at 512 points inside the unit cell of each of the three primitive cubic lattices.

VIII. DIPOLAR ANISOTROPY IN ANTIFERROMAGNETIC MnO

Recently Kaplan[10] has calculated the dipolar anisotropy energy in the antiferromagnet MnO by use of dipole-field factors at various sites within an fcc lattice. We show here how this energy may be obtained in a very simple way by use of the dipole-wave sums.

Consider the fcc sum $D^{ij}(4,4,4)$. The exponential factor of this sum is just ±1 at the various lattice points of an fcc, and in particular reproduces an antiferromagnetic array such that all values $+1$ fall on alternate (111) planes, and the values -1 on the in-between planes. Just such an array is observed in MnO by neutron diffraction experiments.[11] The dipolar energy in such an array of N spins is given by

$$E=-\tfrac{1}{2}N(g\beta s)^2D^{zz}(4,4,4). \quad (27)$$

Here $D^{zz}(4,4,4)$ is the dipole-wave sum with respect to the spin direction. If this latter has direction cosines α,β,γ with respect to the cube axes, we have in terms of our tabulated dipole-wave sums:

$$E=-\tfrac{1}{2}N(g\beta s)^2(\alpha^2D^{xx}+\beta^2D^{yy}+\gamma^2D^{zz}+2\alpha\beta D^{xy}+2\alpha\gamma D^{xz}+2\beta\gamma D^{yz}), \quad (28)$$

where all the D^{ij} have the argument (4,4,4). From Table V we find

$$D^{ij}(4,4,4)=-3.615(1-\delta^{ij}),$$

and thus

$$E=3.615N(g\beta s)^2(\alpha\beta+\alpha\gamma+\beta\gamma). \quad (29)$$

Kaplan obtains the same result, his numerical constant being $(57.41/16)=3.588$.

ACKNOWLEDGMENTS

The authors wish to thank Westinghouse Research Laboratories for their hospitality during the summer of 1953 when the major portion of the computational work of this paper was performed.

[9] J. M. Luttinger and L. Tisza, Phys. Rev. **70**, 954 (1946); Phys. Rev. **72**, 257(E) (1947).

[10] J. I. Kaplan, J. Chem. Phys. **22**, 1709 (1954).

[11] Shull, Strauser, and Wollan, Phys. Rev. **83**, 333 (1951).

Dipolar Ferromagnetism at 0°K

M. H. COHEN, *Institute for the Study of Metals, University of Chicago, Chicago, Illinois*

AND

F. KEFFER, *Sarah Mellon Scaife Radiation Laboratory, University of Pittsburgh,* Pittsburgh, Pennsylvania*

(Received March 11, 1955)

A critical examination of the Holstein-Primakoff spin-wave technique has yielded explicit criteria for the stability of ferromagnetic arrays. These criteria have been applied to point dipoles on the three primitive cubic lattices for the shape most favorable to ferromagnetism, that of a long, thin sample. The results are: (1) dipolar ferromagnetism cannot occur in the simple cubic lattice; (2) the ferromagnetic state can be the lowest, or at least a metastable, state on the face-centered cubic and body-centered cubic lattices. The zero-point energies of ferromagnetic states are calculated for the ⟨100⟩, ⟨011⟩, and ⟨111⟩, directions, yielding two anisotropy constants, K_1 and K_2, for each lattice. The Holstein-Primakoff Hamiltonian is shown not to be diagonal for spin waves in the pathological region, i.e., of wavelengths comparable to the dimensions of the specimens. The properties of this region are shown to be unimportant in energy considerations but may effect stability. They are of great importance in effects such as ferromagnetic resonance which involve long-wavelength spin waves.

I. INTRODUCTION

THE question of the possibility of magnetic order in a lattice of point dipoles is one that has been of interest for a number of years and has been discussed from widely varying points of view. In a high-temperature approximation, Onsager[1] has treated the problem classically and Van Vleck[2] quantum mechanically. Both found that a Lorentz "$4\pi/3$ catastrophe" should not occur and that dipolar ferromagnetism is hence impossible. Classical considerations at 0°K by Sauer[3] and by Luttinger and Tisza[4] indicated that for a simple cubic (sc) lattice an antiferromagnetic array has the lowest energy for all specimen shapes. On the other hand, ferromagnetism was found to be possible in the body-centered cube (bcc) and in the face-centered cube (fcc) for ellipsoids of revolution with sufficiently high axial ratios. Recently Lax[5] has studied the sc lattice over a wide range of temperatures by means of the classical "spherical approximation" and found a transition into an antiferromagnetic state at low temperatures. Kittel[6] has argued from the point of view of domain theory that, if ferromagnetism is possible, it should occur for all specimen sizes and shapes. Tessman[7] has calculated approximately the anisotropy constant K_1 in a ferromagnetic array assumed stable at 0°K.

In this paper, we examine the quantum-mechanical conditions of stability and calculate the zero-point energy of ferromagnetic arrays of point dipoles in the three primitive cubic lattices. The method used is the spin-wave approximation of Bloch, as modified to include dipole-dipole interactions by Holstein and Primakoff (HP).[8] In Sec. II, criteria for the stability of a ferromagnetic array are derived. In Secs. III and IV, we apply these criteria numerically to a representative set of points in **k**-space, using the dipole-wave sums calculated in the previous paper (CK I).[9] The result is that a sc lattice cannot support dipolar ferromagnetism, whereas a fcc or bcc long, thin cylinder can be ferromagnetic. Numerical values are given for the zero-point energies and anisotropy constants K_1 and K_2. It is found that the quantum-mechanical energy of the ferromagnetic state is sufficiently below the classical energies of other arrays to suggest that ferromagnetism will occur for spins greater than 5/2 in the fcc and 6 in the bcc. Finally, we show in Sec. V that the HP approximation does not diagonalize the Hamiltonian for spin waves of wavelengths comparable to the sample dimensions. Although these give unimportant contributions to the energy, they may affect stability. They certainly play an important role even when exchange is present, as in ferromagnetic resonance.

II. STABILITY CRITERIA

The standard ferromagnetic spin-wave approximation including dipole-dipole interactions has been given by Holstein and Primakoff.[8] We shall consider a specimen of unit volume and use only the dipole-dipole portion of the HP Hamiltonian:

$$\mathcal{H}=\tfrac{1}{2}\sum_{l,m}{}'(g^2\beta^2/R_{lm}^5)\times[R_{lm}^2(\mathbf{s}_l\cdot\mathbf{s}_m)-3(\mathbf{s}_l\cdot\mathbf{R}_{lm})(\mathbf{s}_m\cdot\mathbf{R}_{lm})]. \quad (1)$$

As HP show, in the spin-wave approximation this Hamiltonian may be brought into the form

$$\mathcal{H}=C+\sum_{\mathbf{k}}A_{\mathbf{k}}a_{\mathbf{k}}^*a_{\mathbf{k}}+\tfrac{1}{2}\sum_{\mathbf{k}}(B_{\mathbf{k}}a_{\mathbf{k}}a_{-\mathbf{k}}+B_{\mathbf{k}}^*a_{\mathbf{k}}^*a_{-\mathbf{k}}^*), \quad (2)$$

* Supported by the U. S. Air Force under a contract monitored by the Office of Scientific Research of the Air Research and Development Command.

[1] L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936).
[2] J. H. Van Vleck, J. Chem. Phys. **5**, 320 (1937).
[3] J. A. Sauer, Phys. Rev. **57**, 142 (1940).
[4] J. M. Luttinger and L. Tisza, Phys. Rev. **70**, 954 (1946); Phys. Rev. **72**, 257(E) (1947).
[5] M. Lax, J. Chem. Phys. **20**, 1351 (1952).
[6] C. Kittel, Phys. Rev. **82**, 965 (1951).
[7] J. Tessman, Phys. Rev. **96**, 1192 (1954).
[8] T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).
[9] M. H. Cohen and F. Keffer, preceding paper [Phys. Rev. **99**, 1128 (1955)].

where a_k^* and a_k are spin-wave creation and destruction operators, and

$$C=-\tfrac{1}{2}\rho\sum_l'(g^2\beta^2s^2/R_l^5)(3Z_l^2-R_l^2);$$

$$A_k=\sum_l'(g^2\beta^2s/R_l^5)(3Z_l^2-R_l^2)+\tfrac{1}{2}\sum_l'(g^2\beta^2s/R_l^5)(3Z_l^2-R_l^2)e^{i\mathbf{k}\cdot\mathbf{R}_l};$$

$$B_k=-\tfrac{3}{2}\sum_l'(g^2\beta^2s/R_l^5)(X_l^2-Y_l^2-2iX_lY_l)e^{i\mathbf{k}\cdot\mathbf{R}_l}.$$

In the foregoing, ρ is the number of lattice points in the unit volume, and the Z-axis is taken along the direction of magnetization. Throughout the paper we shall use capital letters to signify the axes of quantization and lower case to signify the cubic axes. At this point, it has already been assumed by HP that sums of the form

$$\sum_m' f(\mathbf{R}_{lm})e^{i\mathbf{k}\cdot\mathbf{R}_{lm}}$$

are independent of l except for sites near the surface of the specimen. We shall proceed for the present as if this assumption were true and discuss in Sec. V its failure for the pathological region of $\mathbf{k}$-space defined in CK I, Sec. V.B.

In terms of the notation introduced in CK I we may write

$$C=-\tfrac{1}{2}M_0^2D^{ZZ}(0), \tag{3a}$$

$$A_k=\tfrac{1}{2}(M_0^2/\rho s)[2D^{ZZ}(0)+D^{ZZ}(\mathbf{k})], \tag{3b}$$

$$B_k=-\tfrac{1}{2}(M_0^2/\rho s)[D^{XX}(\mathbf{k})-D^{YY}(\mathbf{k})-2iD^{XY}(\mathbf{k})]; \tag{3c}$$

where $M_0=\rho g\beta s$ is the Z-component of the magnetization when all the spins are perfectly aligned. Here the dipole-wave sums are defined with respect to the axes of quantization X, Y, Z (which we call x^I). Since these sums are given in CK I with respect to the cubic axes x, y, z (which we call x^i) it is necessary to effect a transformation. Defining direction cosines α_{Ii} such that $x^I=\sum\alpha_{Ii}x^i$, we have

$$D^{IJ}=\sum_{ij}\alpha_{Ii}\alpha_{Jj}D^{ij}.$$

This transformation introduces such complexities into the expression for the spin-wave energy for *arbitrary* direction of magnetization that we have found it expedient to consider only the ⟨100⟩, ⟨110⟩, and ⟨111⟩ directions. The explicit numerical results for these three directions are assumed representative of the general character of the anisotropy of the energy.

We may bring Eq. (2) into diagonal form by transformation similar to those used by HP. Since HP included strong exchange interactions, however, they were not concerned with the possibility that the spin-wave approximation might break down, i.e., that the ferromagnetic array might not be stable. With only dipole-dipole interaction present the possibility arises of an unstable aligned array. Accordingly the transformation of Eq. (2) to diagonal form must be done with sufficient care to yield explicit criteria for the stability of the ferromagnetic array within the spin-wave approximation.

We first make the transformation

$$a_{\pm k}=2^{-\frac{1}{2}}(b_k\pm b_{-k}), \tag{4}$$

where $\mathbf{k}$ and $-\mathbf{k}$ are understood to be limited respectively to the regions of the first BZ above and below the k_Xk_Y plane. This transformation unscrambles $\mathbf{k}$ and $-\mathbf{k}$ in Eq. (2), bringing the latter into the form

$$\mathcal{H}=\sum_k{}^+[A_k(b_k^*b_k+b_{-k}^*b_{-k})+\tfrac{1}{2}B_k(b_kb_k-b_{-k}b_{-k})+\tfrac{1}{2}B_k^*(b_k^*b_k^*-b_{-k}^*b_{-k}^*)]+C, \tag{5}$$

where $\sum^+$ means summation over that half of the first BZ above the k_Xk_Y plane. The transformation in Eq. (4) preserves the commutation property of the spin-wave operators.

A transformation diagonalizing the Hamiltonian is

$$b_{\pm k}=l_1c_{\pm k}\pm l_2c_{\pm k}^*, \tag{6}$$

which will preserve the commutation property if

$$|l_1|^2-|l_2|^2=1. \tag{7}$$

The diagonalization of the Hamiltonian by direct substitution of (6) into (5) yields explicit equations for l_1 and l_2. These can be solved under the restriction (7) if and only if, for all $\mathbf{k}$,

$$|B_k|<|A_k|. \tag{8}$$

Equation (8) is the first criterion for the existence of the ferromagnetic spin-wave approximation. Even the case $|B_k|=|A_k|$ must be excluded.

The Hamiltonian in diagonal form becomes

$$\mathcal{H}=C-\tfrac{1}{2}\sum_kA_k+\sum_kE_k(c_k^*c_k+\tfrac{1}{2}), \tag{9}$$

where

$$E_k=E_{-k}=\pm(|A_k|^2-|B_k|^2)^{\frac{1}{2}} \tag{10}$$

and the summation is extended over the entire BZ. The sign of the spin-wave energy E_k is that of A_k, a real quantity. If any A_k is negative the total energy may be lowered by the excitation of spin waves. Thus, we have as the second criterion for stability that for all $\mathbf{k}$

$$A_k>0. \tag{11}$$

The two stability criteria (8) and (11) seem intuitively obvious upon inspection of Eq. (9), an equation given by HP. However, the step between anticipation by inspection and rigorous deduction is not trivial, and we need not apologize for having redone the well-known HP transformation.

We believe the physical interpretation of the stability criteria to be as follows: It can be seen from the derivation of Eq. (3) that the A_k represent the interaction energy between the local magnetic field and the Z-components of the spin-deviations associated with a spin wave. Similarly the B_k represent the interaction energy of the lateral components of the spin deviations with their own lateral magnetic field. Thus a positive A_k is required for the suppression of longitudinal spin deviation. On the other hand, $|A_k|^2$ must be greater

than $|B_{\mathbf{k}}|^2$ in order that the energy increase required by the longitudinal spin deviations exceed the lowering of energy by the self-interaction of the concomitant transverse deviations.

The quantum-mechanical problem of dipolar ferromagnetism is now resolved into two parts:

(1) The detailed testing of $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$ for all $\mathbf{k}$ according to the stability criteria (8) and (11).

(2) With stability assured, the calculation of the total energy and its anisotropy.

We restrict ourselves to $T=0$°K and deal only with the ground-state energy. Since we are concerned primarily with the question of the existence of dipolar ferromagnetism, we further restrict ourselves to consideration of the most favorable specimen shape, a long, thin cylinder. It is perhaps possible that arrays other than the ferromagnetic, e.g., an antiferromagnetic, also are stable and even of lower energy. Since spin-wave treatments of antiferromagnetic arrays are less accurate and more complex than those of ferromagnetic ones we shall not consider them further here. Rather we shall look to the possibility that the quantum-mechanical zero-point energy of a ferromagnetic array is sufficiently less than the difference between the classical energy of it and any antiferromagnetic array to permit identification of the most stable array by classical considerations alone.

It is further possible that the ferromagnetic state be the most stable in *any* shape, since as Kittel[6] has pointed out, dipolar domains can always be formed to eliminate the demagnetizing effects.

Observing that

$$\sum_{\mathbf{k}} D^{IJ}(\mathbf{k})=0, \tag{13}$$

we may sum the second term in Eq. (9) directly and obtain for the ferromagnetic ground-state energy

$$W_0=-\tfrac{1}{2}D^{zz}(0)M_0^2+\tfrac{1}{2}[\textstyle\sum_{\mathbf{k}}E_{\mathbf{k}}-D^{zz}(0)M_0^2/s]. \tag{14}$$

Since $D^{zz}(0)=4\pi/3$, the dipole-field factor for a long, thin cylinder, it is clear that the first term in (14) is the classical energy and that the remaining part of W_0 is the quantum-mechanical zero-point energy.

III. ZERO-POINT SPIN-WAVE ENERGY AND ANISOTROPY CONSTANTS

We have not found it possible to study the stability criteria or to calculate the zero-point energy by analytic methods. Instead we choose a set of 512 points in $\mathbf{k}$-space at which to test the stability criteria and then use values of $E_{\mathbf{k}}$ at these points constructed from the tables in CK I to obtain the zero-point energy by numerical integration. A more detailed discussion of the choice of points in $\mathbf{k}$-space and the numerical integration technique is given in CK I. As pointed out there, the dipole-wave sums are well-behaved except in the vicinity of the origin in $\mathbf{k}$-space. Consequently, the tests of stability in the calculation of zero-point energy are straightforward. We summarize the results of that numerical work in this section, postponing to Secs. IV and V a careful study of spin waves of small wave number.

TABLE I. Zero-point energies and anisotropy constants.

		fcc	bcc
U_0	100	1.5297	1.5475
U_0	110	1.5239	1.4989
U_0	111	1.5072	1.4934
K_0		1.5297	1.5475
K_1		−0.0235	−0.1939
K_2		−0.3966	+0.2856

Explicit calculations were made for a long, thin cylinder, magnetization parallel to the axis of the cylinder. The axis was given in turn the ⟨100⟩, ⟨110⟩, and ⟨111⟩ directions in the three primitive cubic lattices.

The ferromagnetic array is found to be unstable for all three directions of magnetization in the sc. Both stability criteria were found to break down. This fact apparently is related to the fact that of the eight arrays studied classically by Luttinger and Tisza (LT),[4] three were found to have lower energies than the parallel array. As pointed in CK I (see Table VII), any LT array for the sc can be described by a single wave number. Thus a spin wave of wave number coinciding with that of a low energy array should produce a strong magnetic field favoring disalignment. It was actually found that many more spin waves were unstable.

The fcc and bcc ferromagnetic arrays both prove to be stable for all three directions of magnetization. Using our interpretation of the meaning of the stability criteria, we are again in agreement with the classical results of LT.

As shown in Sec. IV, the contribution of the cell about the origin to the zero-point energy is negligibly small. We present in Table I the results of our numerical calculations neglecting the contribution of the first cell. In Table I, we have put

$$U_0=\tfrac{1}{2}(s/M_0^2)\textstyle\sum_{\mathbf{k}}E_{\mathbf{k}}. \tag{15}$$

Since calculations were made for three different orientations we can determine two anisotropy constants. To do so we approximate the energy by

$$W_0=-\tfrac{2}{3}\pi M_0^2+(M_0^2/s)\{-(\tfrac{2}{3}\pi)+K_0 +K_1(\alpha_1^2\alpha_2^2+\alpha_1^2\alpha_3^2+\alpha_2^2\alpha_3^2)+K_2\alpha_1^2\alpha_2^2\alpha_3^2\}, \tag{16}$$

where $\alpha_1, \alpha_2, \alpha_3$ are the direction cosines of the magnetization with respect to the cubic axes. Values of K_0, K_1, and K_2 are listed in Table I. Within the approximation (16) the ⟨111⟩ orientation yields the lowest energy of any orientation for both lattices. It is clear that K_1 alone does not suffice to describe the orientational dependence of the energy.

To decide rigorously about the absolute stability of the ferromagnetic array, we would have to make

calculations of the zero-point energy of all other types of arrays. It seems reasonable, however, that the ratio of zero-point energy to classical energy should be roughly comparable for all arays. According to LT, in the fcc and bcc the classical energy of the long, thin ferromagnet is lower than the energy of the lowest antiferromagnetic state considered by them by $0.286M_0^2$ and $0.108M_0^2$, respectively. If the zero-point energy of the ferromagnetic array,

$$(M_0^2/s)[U_0-(2\pi/3)],$$

is small compared to this classical difference, we can perhaps conclude that the ferromagnetic array is absolutely stable. The lowest zero-point energy ($\langle 111\rangle$ direction) is $0.587(M_0^2/s)$ and $0.601(M_0^2/s)$ for the fcc and bcc respectively. For us to be secure in the conclusion that the ferromagnetic state is absolutely stable, the spin s would have to be greater than 5/2 (fcc) or 6 (bcc).

It is of interest to check the accuracy of the numerical integration by use of the relation

$$\sum_{\mathbf{k}}|B_{\mathbf{k}}|^2=\frac{9}{4}\frac{M_0^4}{\rho s^2}\times\frac{1}{\rho^2}\sum_l{}'\frac{X_l^2+Y_l^2}{R_l^{10}}. \tag{17}$$

The lattice sum in (17) is expressed with respect to the axes of quantization. When it is expressed with respect to the cube axes, Eq. (17) becomes

$$\sum_{\mathbf{k}}|B_{\mathbf{k}}|^2=\frac{9}{4}\frac{M_0^4}{\rho s^2}(\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_1^2\alpha_3^2)\times\frac{1}{\rho^2}\sum_l{}'\frac{r_l^4-5z_l^4}{r_l^{10}}. \tag{18}$$

The left-hand side of (18) was evaluated by numerical integration; the right-hand side by direct summation. The agreement of the two results was within 2%.

If a large magnetic field H_0 is applied in the direction of quantization, it becomes possible to expand $E_{\mathbf{k}}$ in powers of $|B_{\mathbf{k}}|^2/g\beta H_0$. In this approximation, as can be seen from (18), the first term in the expansion yields an anisotropy constant K_1.

Recently Tessman[7] has calculated K_1 by assuming that the expansion can be done in the absence of the field, that is, by replacing all $A_{\mathbf{k}}$ by A_0 and expanding in powers of $|B_{\mathbf{k}}|^2/A_0$. However, our detailed numerical study shows that $A_{\mathbf{k}}$ depends strongly on $\mathbf{k}$ and that the magnitudes of $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$ are comparable over most of $\mathbf{k}$-space. Nevertheless Tessman's calculation does yield a reasonably good value of K_1 for the bcc. The test of our numerical work based on Eq. (18) has incidentally checked accurately the results of Tessman's expansion for both bcc and fcc.

In a discussion of domain formation in dipolar ferromagnetism, Kittel[6] has estimated the anisotropy constant K_1 to be approximately M_0^2/s. This is an overestimate, as shown by Table I, and further K_1 alone does not suffice to describe the anisotropy energy. In fact, from the size of K_2 it would appear that even K_1 and K_2 together might not be adequate to describe the anisotropy energy. However, the general conclusions reached by Kittel remain correct. In particular, as discussed further in Sec. V, if ferromagnetism is possible in the long, thin cylinder, it is possible in specimens of any shape. It should therefore be pointed out that in experiments with presumed dipolar arrays in an applied magnetic field, care must be taken not to confuse (A) the transition from possible antiferromagnetism with (B) the possible domain alignment of ferromagnetism.

IV. STABILITY AND ZERO-POINT ENERGY OF SPIN WAVES IN THE FIRST CELL

The results of CK I concerning the behavior of dipole-wave sums near $\mathbf{k}=0$ indicate that the stability and energy of the $\mathbf{k}=0$ spin wave are not representative of the cell about $\mathbf{k}=0$. In this section, we study those spin waves whose wave numbers lie outside the pathological region about $\mathbf{k}=0$ but still within the first cell.

The lowest order terms in $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$ depend upon the direction of $\mathbf{k}$ only and do not satisfy the stability criterion (8). The integral approximation to $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$ developed by HP for arbitrary $\mathbf{k}$ in a long, thin cylinder gives in fact the lowest order terms obtained in the limit of small $\mathbf{k}$ for the infinite lattice. This instability of spin waves of small $\mathbf{k}$ thus, in effect, was first pointed out by HP. However, by including the structure dependent terms of order k^2 [10] in $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$, we have been able to show that the stability criteria *are* satisfied for all three directions of magnetization in both fcc and bcc. When the direction of the spin-wave wave number coincides with the axis of magnetization, it is necessary to go to order k^4 in $E_{\mathbf{k}}^2$ to prove it >0. These considerations are tedious, detailed, and entirely numerical and will not be given here. As an example of the dispersion relation in this region we give $E_{\mathbf{k}}$ for the fcc magnetized along a $\langle 100\rangle$ direction:

$$E_{\mathbf{k}}\langle 100\rangle=2.446(M_0^2/\rho s)ka\sin\theta_{\mathbf{k}}\times[1-0.884\sin^2\theta_{\mathbf{k}}\sin^2 2\phi_{\mathbf{k}}]^{\frac{1}{2}}. \tag{19}$$

A similar result obtains for the bcc for the $\langle 100\rangle$ direction, the constants being 3.054 and 0.866, respectively. More complicated expressions obtain for other directions of magnetization.

We replace the first cell by a sphere of the same volume and average Eq. (19) for the spin-wave energy over the sphere. This approximate average spin-wave energy in the first cell turns out to be of order 10^{-4} with respect to the spin-wave energy averaged over the entire BZ. Hence the neglect of the first cell in the calculation of Sec. III is justified.

[10] See Eq. (18') of CK I.

V. SPIN WAVES IN THE PATHOLOGICAL REGION

As can be seen by tracing through the derivation of Eq. (2), the form of the Hamiltonian expressed therein depends upon the fact that sums of the form $\sum_m f(\mathbf{R}_{lm})e^{i\mathbf{k}\cdot\mathbf{R}_{lm}}$ are independent of l except for lattice points very near the surface. If there are wave numbers for which these sums are not independent of l, there will be a coupling among all spin waves of those wave numbers. The Hamiltonian will not be diagonal, and a complete discussion of the ground-state energy of a ferromagnetic array will not be possible within the framework of the HP approximation.

We have evaluated a dipole-wave sum typical of those entering $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$ as an explicit function of the point of an infinite cylinder for $\mathbf{k}$'s in the pathological region. The result is

$$D^{zz}(l,\mathbf{k})=\frac{1}{\rho}\sum_m{}'\frac{3Z_{lm}^2-R_{lm}^2}{R_{lm}^5}e^{i\mathbf{k}\cdot\mathbf{r}_{lm}}=\frac{4\pi}{3}\left(1-\frac{3k_Z^2}{k^2}\right)$$

$$+\frac{4\pi k_Z^2}{k^2}e^{i\mathbf{k}_1\cdot\boldsymbol{\eta}_l}\sum_{n=-\infty}^{\infty}(-i)^n I_n(k_Z\eta_l)[k_1LK_n(k_ZL)J_{n-1}(k_1L)$$

$$+k_ZLK_{n-1}(k_ZL)J_n(k_1L)], \quad (20)$$

where k_Z and $\mathbf{k}_1$ are components of $\mathbf{k}$ parallel and perpendicular to the axis of the cylinder, $\boldsymbol{\eta}_l$ is the radial part of the lattice vector $\mathbf{R}_l$, and L is the radius of the cylinder. For $kL\lesssim 10$, i.e., in the pathological region, Eq. (20) is strongly dependent upon $\boldsymbol{\eta}_l$ for all l. In the special case $\mathbf{k}=0$, however, it does go to the correct limit $4\pi/3$, independent of l. It should be noted, therefore, that there is no coupling between $\mathbf{k}=0$ spin waves and others.

The coupling among $\mathbf{k}\neq 0$ spin waves has its origin in the introduction of spin-wave operators by HP through the transformation

$$a_l=N^{-\frac{1}{2}}\sum_{\mathbf{k}}a_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{R}_l}. \quad (21)$$

We can thus ask the question, is it possible to find a different transformation to spin-wave operators, one presumably more appropriate to the long, thin cylinder than the one involving plane waves and one which does diagonalize the Hamiltonian? We have not been able to find such a transformation. The best we have been able to do is to simplify matters by the introduction of the cylindrical waves $e^{i(hZ_l+\nu\Phi_l)}J_\nu(\alpha\eta_l)$. We choose α so that $J_\nu(\alpha L)=0$. The question posed above then becomes: are

$$\sum_{lm}{}'e^{-i(h'Z_l+\nu'\Phi_l)}J_{\nu'}(\alpha'\eta_l)\frac{3Z_{lm}^2-R_{lm}^2}{R_{lm}^5}$$

$$\times e^{i(hZ_m+\nu\Phi_m)}J_\nu(\alpha\eta_m) \quad (22)$$

and similar forms now entering the Hamiltonian diagonal in $h'\nu'\alpha'$ and $h\nu\alpha$?

The technique used to evaluate the expression (22) is, as usual, to perform an infinite lattice sum over m and subtract from it the integral over space outside the cylinder. Then we replace the summation over l by an integral over the cylinder. The artifice of the infinite sum is introduced because of the exclusion of the point l in the summation over m. Actually the infinite sum is used only in principle, it being sufficient to show that the term resulting from the integral outside the cylinder is nondiagonal. The result for this latter term of (22) is

$$-\frac{2(2\pi)^3h^2\alpha\alpha'L^2}{(h^2+\alpha^2)(h'^2+\alpha'^2)}J_{\nu-1}(\alpha L)J_{\nu-1}(\alpha' L)$$

$$\times K_\nu(hL)I_\nu(hL)\delta(h-h')\delta_{\nu\nu'}. \quad (23)$$

Again there is no coupling of $\mathbf{k}=0$ spin waves to others.

Thus, since we are unable to diagonalize the Hamiltonian in this small pathological region, we cannot conclude rigorously that the ferromagnetic state is stable. If it should turn out, however, that appropriate stability criteria are satisfied by the unknown form of the diagonalized Hamiltonian, the values for the zero-point energy and anisotropy constants given in Sec. III would remain unaltered.

For arbitrary shaped specimens, Kittel's[6] ideas about dipolar ferromagnetic domains can be studied within the spin-wave approximation. To do so one quantizes each spin about the orientation of the magnetization in the analogous classical domain structure. One can then introduce spin-deviation operators and assume their expectation values to be small. If one now uses a spin-wave transformation mixing only spins in a given domain, one can use the HP technique to diagonalize the Hamiltonian. For spin waves of wavelengths comparable to the dimensions of the domains there will be interdomain interactions. But these are spin waves in the pathological region and interact within a domain. If they do not violate the stability criteria for a single domain, they probably will not violate them for many domains, and cannot contribute appreciably to the energy of the entire system. Thus it seems reasonable to conclude with Kittel that if dipolar ferromagnetism is possible in any one shape it is possible in all shapes.

It should be strongly emphasized that this pathological behavior at small $\mathbf{k}$ is not peculiar to spin waves in dipolar ferromagnetism alone. In an ordinary ferromagnet the exchange contribution is completely negligible at these wave numbers, and the behavior of the Hamiltonian is dominated by the long-range dipole terms. Thus physical phenomena strongly dependent upon small wave-number spin waves, such as ferromagnetic resonance or thermal behavior below 1°K, must be affected in detail by the pathological behavior of the long-range dipolar interaction in a finite sample. In particular, the problem of the line width caused by the skin-depth effect in ferromagnetic resonance is

completely unsettled. Even if the skin depth is comparable to the dimension of the specimen, there will be coupling between the microwave magnetic field and all of the spin waves in the pathological region. No technique for calculating the skin-depth width has occurred to us which would not require the diagonalization of the Hamiltonian. We have in fact verified that in a case of practical importance in metallic ferromagnetic resonance, i.e., the flat-plate specimen, position dependence does also occur in the dipole-wave sums.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Electron-Phonon Interaction in Metals*

JOHN BARDEEN AND DAVID PINES†
Physics Department, University of Illinois, Urbana, Illinois
(Received April 4, 1955)

The role of electron-electron interactions in determining the electron-phonon interaction in metals is investigated by extending the Bohm-Pines collective description to take into account the ionic motion. Collective coordinates are introduced to describe the long-range electron-ionic correlations, and it is shown by a series of canonical transformations that these give rise to plasma waves and to coupled electron-ion waves which correspond to longitudinal sound waves. The dispersion relation for the sound waves is identical with that derived by Toya and Nakajima by self-consistent field methods. The velocity of these sound waves is calculated from first principles for sodium and is found to be in good agreement with experiment. The effective matrix element for the electron-phonon interaction is determined and is found to be identical for long wavelengths with that found earlier by Bardeen using a self-consistent field method which neglects exchange and correlation effects. The agreement with the earlier work is explained by the fact that the residual electron-electron interaction is of quite short range, so that an independent-particle treatment is rather well justified. The effects of Coulomb correlations on superconductivity are likewise shown to be small, so that the neglect of Coulomb interactions in the formulation of the superconductivity problem is justified.

1. INTRODUCTION

IN the coupled motion of electrons and ions in a metal, the tendency of the electrons to screen out the field of the ions is of primary importance. This screening greatly influences the effective matrix element for the electrical conductivity of the metal; it is also determining for the longitudinal sound wave velocity. The extent and effectiveness of this screening depend, in turn, on the electron-electron interaction in the metal. The usual theories of metallic electrical conduction treat these electron interactions on an independent-particle model in which each electron is assumed to move independently in a self-consistent field determined by the ions and the other electrons. Often no attempt is made to determine this self-consistent field; instead an empirical constant is introduced to describe the electron-lattice interaction. In 1937, one of us extended the self-consistent field method to take into account the motion of the ions, and thus to determine the matrix element for electron-lattice interactions.[1] The electron-electron interactions were treated in the Hartree approximation. More recently Nakajima[2] has given a simple field-theoretic treatment of this problem which also invokes the Hartree approximation and which leads to results essentially equivalent to *A*.

The independent-particle model has been successfully applied to a wide variety of problems. This success is somewhat surprising, because the marked correlations in electronic positions due to Coulomb interaction and exchange have been neglected. One of the main objects of this paper is to give a justification for this approach to the theory of conductivity.

The question of electronic correlations is of particular importance in the theory of superconductivity. Here one would expect that the large Coulomb correlation energy would play a more important role than the relatively small electron-lattice interaction energy in determining the transition from the normal to the superconducting state. However, experimentally this is not the case, as is shown by the isotope effect. We shall show why this is to be expected on theoretical grounds, and thereby justify the application of the independent-particle model to the formulation of this problem.

The physical basis for the unimportance of electron-electron correlations in conductivity and superconductivity arises from the tendency of the electrons to stay apart from one another in such a way that the field of a given electron is screened out within a distance of the order of the inter-particle spacing. Bohm and one of the authors have established this by showing that the long-range part of the Coulomb interaction leads only to coherent plasma oscillations of the electron gas and may be described in terms of these oscillations.[3] Energies of the plasma quanta are so high that the

* This research was supported in part by the Office of Ordnance Research, U. S. Army.
† Now at: Palmer Physical Laboratory, Princeton University, Princeton, New Jersey.
[1] J. Bardeen, Phys. Rev. **52**, 688 (1937), hereafter referred to as A.
[2] S. Nakajima, Proceedings of the International Conference on Theoretical Physics, Kyoto and Tokyo, September, 1953 (Science Council of Japan, Tokyo, 1954).
[3] D. Bohm and D. Pines, Phys. Rev. **82**, 625 (1951); **85**, 338 (1952); **92**, 609 (1953); and D. Pines, Phys. Rev. **92**, 626 (1953), hereafter referred to as BP I, BP II, BP III, and P IV respectively.

oscillations are not excited thermally. There remains the short-range screened interaction between the aforementioned electrons. This analysis shows why electron-electron collisions are not important for either thermal or electrical conduction in metals. Abrahams[4] has estimated the collision cross section and mean free path for electrons interacting via the screened Coulomb interaction in alkali metals. While the cross sections are of a reasonable order of magnitude ($\sim 10^{-15}$ cm^2) the scatterings possible are so greatly restricted by the exclusion principle that collisions are infrequent. The electron-electron mean free path is long compared with the electron-phonon mean free path at all temperatures of interest.

The BP collective description is extended here to treat the coupled system of moving electrons and ions. Plasma coordinates are introduced to take into account the effect of electron-electron interactions on the electron-ion system. It is found that the effect of the long-range interactions may then be described rather simply, and explicit expressions are derived for the effective electron-phonon matrix element and for the sound wave frequencies. There remains the short-range electron-electron interaction which, as we have mentioned, does not influence the motion of a given electron appreciably, so that it is not surprising that superconductivity is little affected by the electronic Coulomb correlations.

The results we obtain for the influence of the electronic response on the electron-phonon interaction are in good agreement with those obtained in A. This problem has also been treated from another point of view by Bohm and Staver,[5] who describe the ions and electrons as a set of coupled plasmas. They used a semiclassical approach closely related to BP II, and obtained the phonon dispersion relation. Our dispersion relation reduces to theirs in the appropriate limit.

The foregoing collective description is not applicable in the short wavelength region (corresponding to Fourier components of the electron-electron or electron-phonon interaction of wave vector greater than about the inverse particle spacing). Here the electron response to the ionic motion is probably best accounted for using the formulation of Nakajima. However, this formulation should be extended to include exchange effects, for it is precisely in this region (distances comparable or small compared to an electron de Broglie wavelength) that exchange begins to play an important role. We discuss the Nakajima method with exchange, but do not attempt to apply it to a detailed calculation because of the mathematical difficulties encountered in so doing.

2. DERIVATION OF HAMILTONIAN

We shall derive here for a monatomic crystal the Hamiltonian which will be used in the subsequent discussion. A number of approximations are made in order to simplify the calculations. It is assumed, as is usually done in the Bloch theory, that the matrix elements of the electron-lattice and Coulomb interactions depend only on the difference in wave vectors of the initial and final states. Another approximation is that lattice waves are either longitudinal or transverse, and that the electrons interact only with the longitudinal component. Both of these amount to neglect of anisotropic effects which greatly complicate the equations are probably important only for detailed quantitative calculations.

The general Hamiltonian is of the form

$$H=\sum_i (\mathbf{p}_i^2/2m)+\sum_{i,j} v(\mathbf{x}_i-\mathbf{X}_j) +H_{\text{ion-ion}}+H_{\text{Coul}}. \quad (2.1)$$

where i runs over the valence electrons, j over the ions, $v(\mathbf{x}_i-\mathbf{X}_j)$ represents the electron-ion interaction, $H_{\text{ion-ion}}$ includes the Coulomb and exchange repulsion of the ion cores and H_{Coul} the Coulomb interaction of the electrons. There are $n_0=ZN$ electrons/cm^3 where Z is the valency and N the ionic density. In order to eliminate infinities which appear in the separate terms of (2.1) (although not in the sum) we suppose there is subtracted from the electron-ion interaction the interaction of each electron with a uniform positive charge, from the electron-electron interaction the self-energy of a uniform negative charge, and from the ion-ion interaction the self-energy of a uniform positive charge. The sum of these adds to zero, so that the energy is unchanged. The ion-ion interaction energy less the self-energy of a uniform positive charge is equivalent to the energy of the ions in a uniform negative sea, including the self-energy of the negative charge.

We describe the electron wave function in second quantization by occupation numbers of a set of Bloch functions. It is assumed that the individual electrons move in a potential $V(\mathbf{x})$ which is the potential of the ions compensated by a uniform negative charge:

$$V(\mathbf{x})=\sum_j v(\mathbf{x}-\mathbf{X}_j)+\text{comp. charge}. \quad (2.2)$$

The Bloch equation is

$$[(\mathbf{p}^2/2m)+V(\mathbf{x})]\psi_\kappa=E_\kappa\psi_\kappa. \quad (2.3)$$

We have used the extended zone scheme, so that κ is not necessarily in the first Brillouin zone. A discrete set of κ values is obtained in the usual way by introduction of periodic boundary conditions, with N the number of atoms in a fundamental period of unit volume. Creation and destruction operators, $c_{\kappa s}^*$, $c_{\kappa s}$, are defined in the usual way and obey commutation laws for Fermi particles:

$$[c_{\kappa' s'}^*, c_{\kappa s}]^+=\delta_{\kappa\kappa'}\delta_{ss'}. \quad (2.4)$$

We shall omit the spin index s except where necessary for clarity.

The departure of the ions from equilibrium positions,

$$\delta\mathbf{X}_j=\mathbf{X}_j-\mathbf{X}_j^0, \quad (2.5)$$

[4] E. Abrahams, Phys. Rev. **95**, 839 (1954).
[5] D. Bohm and T. Staver, Phys. Rev. **84**, 836 (1952); T. Staver, Ph.D. thesis, Princeton University, 1952 (unpublished).

may be expressed in terms of normal coordinates, $q_{k\sigma}$ ($\sigma=1, 2, 3$):

$$\delta\mathbf{X}_j=(NM)^{-\frac{1}{2}}\sum_{k\sigma} q_{k\sigma}\boldsymbol{\epsilon}_{k\sigma}\exp[i\mathbf{k}\cdot\mathbf{X}_j^0], \quad (2.6)$$

in which $\boldsymbol{\epsilon}_{k\sigma}$ is a unit vector in the direction of polarization of the lattice waves, taken in the same sense for $\mathbf{k}$ and $-\mathbf{k}$ so that $q_{-k}=q_k^*$. We assume that one of these directions is longitudinal and two transverse, and that the transverse frequencies are determined entirely from ion-ion interactions. These assumptions are not really valid, particularly for short wavelengths for which the distinction between longitudinal and transverse waves is no longer sharp. While much of the analysis can be carried through without making these assumptions, the slight additional generality gained does not seem warranted by the increase in complexity of the equations. We shall use the symbol q_k without polarization subscript to denote the longitudinal mode. In some of the summations, $\mathbf{k}$ may run out of the first zone. In such a case, we always mean the q_k for the corresponding $\mathbf{k}$ in the first zone. The phonon Hamiltonian is

$$H_{\rm ph}=\tfrac{1}{2}\sum_{k\sigma\sigma} (p_k^*p_k+\Omega_k^2q_k^*q_k), \quad (2.7)$$

where Ω_k^2 is determined solely from ion-ion interactions. Calculation of the correct frequency, ω_k, which requires inclusion of the electron-lattice interactions, is discussed later.

The remaining parts of the Hamiltonian are expressed in terms of the c_α's. The electron-lattice interaction may be written

$$\sum_{i,j} v(\mathbf{x}_i-\mathbf{X}_j)=\sum_{i,j} v(\mathbf{x}_i-\mathbf{X}_j^0) - (NM)^{-\frac{1}{2}}\sum_k \boldsymbol{\epsilon}_k\cdot\nabla v(x-X_j^0)q_k\exp[i\mathbf{k}\cdot\mathbf{X}_j^0]. \quad (2.8)$$

The first term may be combined with the kinetic energy of the electrons to give

$$H_{\rm el}=\sum_i[(\mathbf{p}_i^2/2m)+V(\mathbf{x}_i)]=\sum_{\alpha\kappa} E_\alpha c_{\alpha\kappa}^*c_{\alpha\kappa}. \quad (2.9)$$

The second term of (2.8) may be expressed in terms of the matrix elements,

$$v_k{}^i=-(NM)^{-\frac{1}{2}}\int\psi_{\alpha+\kappa}^*\{\sum_j \boldsymbol{\epsilon}_k\cdot\nabla v(\mathbf{x}_i-\mathbf{X}_j^0) \times\exp[i\mathbf{k}\cdot\mathbf{X}_j^0]\}\psi_\alpha d\tau, \quad (2.10)$$

assumed to be independent of $\boldsymbol{\kappa}$, to give

$$H_{\rm int}=\sum_{\alpha\kappa\sigma} c_{\alpha+k,\sigma}{}^*c_{\alpha,\sigma}q_kv_k{}^i\rho_{-k}, \quad (2.11)$$

where

$$\rho_k=\sum_{\alpha\sigma} c_{\alpha-k,\sigma}{}^*c_{\alpha,\sigma}; \quad \rho_{-k}=\sum_{\alpha\sigma} c_{\alpha,\sigma}{}^*c_{\alpha-k,\sigma}. \quad (2.12)$$

The sum over $\mathbf{k}$ and $\boldsymbol{\kappa}$ extend over all values; q_k refers to the reduced vector in the first zone. Note that $(v_k)^*=v_{-k}{}^i$. The Coulomb interaction may be expressed in the form

$$H_{\rm Coul}=\tfrac{1}{2}\sum_k M_k{}^2\rho_{-k}\rho_k, \quad (2.13)$$

where for free electrons, $M_k{}^2=4\pi e^2/k^2$. It has again been assumed that the matrix elements depend only on the vector difference between initial and final states.

The Hamiltonian is now expressed in the form

$$H=E_{\text{ion-ion}}+H_{\rm tr}+H_1,$$

where

$$H_1=H_{\rm el}+H_{\rm ph}+H_{\rm int}+H_{\rm Coul}, \quad (2.14)$$

and $E_{\text{ion-ion}}$ is the ion-ion interaction energy for the ions in equilibrium positions, $H_{\rm tr}$ is the Hamiltonian for transverse phonons, and the terms in H_1 are given by (2.7), (2.9), (2.11), and (2.13). From now on, we shall be concerned with the Hamiltonian H_1. In Sec. III, we discuss earlier methods for calculation of the electron-lattice interaction by self-consistent field methods and in Sec. 4 we introduce plasma coordinates to treat the long-range part of the Coulomb interaction.

3. EARLIER METHODS FOR THE CALCULATION OF v_k AND ω_k

We are concerned in this section with a review of calculations which have been made of the effect of the shielding of the valence electrons on the interaction potential and on the sound-wave vibrational frequencies. For long wavelengths, the interaction of a given electron with the ionic motion is radically altered by the field due to the other electrons, which in the course of responding to the ionic motion move in just such a way as to produce a field which very nearly cancels the ionic field. The effective matrix element for the electron-lattice interaction, v_k, may be written as the sum of $v_k{}^i$ due to the motion of the ions and $v_k{}^e$ due to the motion of the electrons:

$$v_k=v_k{}^i+v_k{}^e. \quad (3.1)$$

In A, $v_k{}^e$ was computed by what amounts to a Hartree self-consistent field calculation. It was assumed that the wave functions of the individual electrons change adiabatically with the ion motion. Recently Toya has determined the longitudinal sound wave frequency, ω_k, using this same model.[6] We here give a simple derivation of the effect of the electron response on v_k and ω_k which agrees in the long wavelength limit with the foregoing calculations. Our treatment parallels in some respects that of Bohm and Staver.

According to (2.14) the equation of motion of the $\mathbf{k}$th sound wave amplitude is

$$\ddot{q}_k+\Omega_k{}^2q_k=-v_{-k}{}^i\rho_k. \quad (3.2)$$

The electronic density fluctuation ρ_k consists of two parts; one, $\rho_k{}^0$, associated with the motion of electrons in the absence of a sound wave, and another, $\delta\rho_k$ which represents the motion of the electrons associated with the sound wave. For pure Coulomb interactions between free electrons $\delta\rho_k$ is related to the Fourier component of that part of the effective one-electron interaction potential v_k associated with the electronic re-

[6] T. Toya, Busseiron Kenkyu **59**, 179 (1952).

sponse v_k^{ρ}, by Poisson's equation; $k^2 v_k^{\rho} q_k = 4\pi e^2 \delta\rho_k$, or more generally,

$$v_k^{\rho} q_k = M_k^{-2} \delta\rho_k. \tag{3.3}$$

Thus we may write for our sound-wave, q_k,

$$\ddot{q}_k + \omega_k^2 q_k = -v_{-k}^{i} \rho_k^{0}, \tag{3.4}$$

with ω_k specified by the dispersion relation

$$\omega_k^2 = \Omega_k^2 + M_k^{-2} v_{-k}^{i} v_k^{\rho}. \tag{3.5}$$

Equation (3.4) describes the interaction of the sound waves with the *independently* moving electrons (here described in terms of their density fluctuation ρ_k^0) and in principle the resistance may be calculated from it.

The matrix element v_k^{ρ} may be simply calculated in the Fermi-Thomas approximation.[7] In this approximation the electron density, $\rho(\mathbf{x})$, is proportional to $[E_0 - \delta V(\mathbf{x})]^{\frac{3}{2}}$, where E_0 is the energy of an electron at the top of the Fermi distribution and $\delta V(\mathbf{x})$ is the effective potential acting on the electrons due to the ionic motion. We then have for the density change $\delta\rho(\mathbf{x})$ associated with the sound-wave potential $\delta V(\mathbf{x})$,

$$\delta\rho(\mathbf{x}) = -\frac{3}{2}\frac{n_0}{E_0}\delta V(\mathbf{x}). \tag{3.6}$$

We Fourier-analyze (3.6) and apply (3.1) to obtain

$$\delta\rho_k = -\frac{3}{2}\frac{n_0}{E_0} v_k q_k = -\frac{3}{2}\frac{n_0}{E_0}(v_k^{i} + v_k^{\rho}) q_k. \tag{3.7}$$

If we now compare (3.7) with (3.3), we see that the requirement that our treatment be self-consistent yields

$$v_k^{\rho} = \frac{-v_k^{i}}{1 + (2E_0/3n_0 M_k^2)}. \tag{3.8}$$

For long wavelengths, M_k^2 is given by its free-electron value, $4\pi e^2/k^2$, and we have

$$v_k^{\rho} = \frac{-v_k^{i}}{1 + (k^2 v_0^2/3\omega_p^2)} \cong -v_k^{i}\left(1 - \frac{k^2 v_0^2}{3\omega_p^2}\right), \tag{3.9}$$

where v_0 is the velocity of an electron at the top of the Fermi distribution and

$$\omega_p^2 = 4\pi n e^2/m \tag{3.10}$$

is the square of the plasma frequency. Thus we see that the effect of the field of the other electrons is to screen out the "bare" electron-ion interaction within a distance of order ω_p/v_0. For long wavelengths the effective matrix element for electron-phonon interaction is thus drastically reduced to

$$v_k \cong (k^2 v_0^2/3\omega_p^2) v_k^{i}. \tag{3.11}$$

The corresponding expression for the sound-wave frequencies is

$$\omega_k^2 = \Omega_k^2 - \frac{k^2}{4\pi e^2} v_{-k}^{i} v_k^{i}\left(1 - \frac{k^2 v_0^2}{3\omega_p^2}\right). \tag{3.12}$$

Our expressions (3.9) for v_0 and (3.12) for ω_k agree in this limit with the corresponding results of A and of Toya.

We may obtain a quite general expression for ω_k if we neglect the effect of lattice periodicity on the Ω_k and v_k^{i} and consider only Coulombic electron-ion and ion-ion interactions. This corresponds to treating the positive ions as a plasma; the "bare" phonon frequency is then

$$\Omega_k^2 \cong \Omega_P^2 = 4\pi N Z^2 e^2/M. \tag{3.13}$$

The v_k^{i} are given by

$$v_k^{i} \cong -(4\pi Z e^2 i/k)(N/M)^{\frac{1}{2}}. \tag{3.14}$$

We then have

$$\omega_k^2 = k^2 v_0^2 \Omega_P^2/3\omega_P^2 = mZ^2 v_0^2 k^2/3M. \tag{3.15}$$

Thus in this approximation the square of the phonon frequency has been reduced by the same factor as the electron-phonon interaction matrix element,[8] the phonon frequency is altered from a constant to one linear in $\mathbf{k}$, with a sound velocity c_l, which depends only on the density and which is given by

$$c_l = mZ v_0/\sqrt{3}M. \tag{3.16}$$

Equation (3.16) was first obtained by Bohm and Staver,[5] and as shown by Staver[5] is in fairly good agreement with the experimental observations. The extent of this agreement ($\sim 20\%$ for the alkali metals, no worse than a factor of two for any metal) indicates that one can get a good order of magnitude estimate, by such simple Coulombic considerations. In Appendix A, we discuss the results obtained with our more accurate dispersion relation (3.12). It is interesting to note that this same simplified model leads to a Sommerfeld-Bethe interaction constant $C = E_F/\sqrt{2}$.

Nakajima's field-theoretic derivation for v_k^{ρ} and ω_k goes somewhat beyond the adiabatic approximation but gives nearly equivalent results. We sketch his derivation here because as we shall see, it offers the best framework for the treatment of the short wavelength electron-phonon interaction. Frohlich[9] has used a field-theoretic method which differs in some respects. He used an individual particle model which did not include Coulomb interactions between the electrons. His choice of creation and destruction operators and method of analyzing is such as to give the correct result for his problem when the change in the frequency from electron-phonon interactions is small. However, his dis-

[7] See also A. W. Overhauser, Phys. Rev. **89**, 689 (1953).

[8] The same shielding factor occurs because in the limit we are considering we may write $\omega^2 = nk^2 v_k^{i}(v_k^{i} + v_k^{\rho})/m$, so that the sound wave frequencies are proportional to the effective matrix element v_k.

[9] H. Frohlich, Proc. Roy. Soc. (London) **A215**, 291 (1952).

persion relation is not correct when the change in frequency is large.

Nakajima writes the Hamiltonian, H_1, in a form equivalent to the following:

$$H_1=\sum_{\kappa s} E_\kappa c_{\kappa s}^* c_{\kappa s}+\tfrac{1}{2}\sum_{\text{some }k}(p_k^* p_k+\omega_k^2 q_k^* q_k) +\sum_k v_k q_k \rho_{-k}+\tfrac{1}{2}\sum_k M_k^2 \rho_k \rho_{-k}+\sum_k (v_k{}^i - v_k) q_k \rho_{-k} +\tfrac{1}{2}\sum_{\text{some }k}(\Omega_k^2-\omega_k^2) q_k^* q_k. \quad (3.17)$$

A canonical transformation is made to eliminate the first linear term in q_k. The interaction v_k is determined so as to eliminate the second linear terms in q_k, the second line, and ω_k^2 is chosen so as to cancel diagonal terms in $q_k^* q_k$ in the last line.

The transformed Hamiltonian is

$$H_1'=e^{-iS/\hbar}H_1 e^{iS/\hbar}=H_1+i\hbar^{-1}[H_1,S] -\tfrac{1}{2}\hbar^{-2}[[H_1,S],S]+\text{etc.} \quad (3.18)$$

For S we take

$$S=i\sum_{k\kappa} c_\kappa^* c_{\kappa-k}[f(\mathbf{k},\boldsymbol{\kappa})q_k - ig(\mathbf{k},\boldsymbol{\kappa})p_{-k}]. \quad (3.19)$$

The required commutators are:

$$[\textstyle\sum_\kappa E_\kappa c_\kappa^* c_\kappa, S]=i\sum (E_\kappa - E_{\kappa-k}) c_\kappa^* c_{\kappa-k} \times[f(\mathbf{k},\boldsymbol{\kappa})q_k - ig(\mathbf{k},\boldsymbol{\kappa})p_{-k}]. \quad (3.20)$$

$$[\tfrac{1}{2}\textstyle\sum (p_k^* p_k+\omega_k^2 q_k^* q_k), S]=i\sum c_\kappa^* c_{\kappa-k} \times\{-i\hbar p_{-k} f(\mathbf{k},\boldsymbol{\kappa})+\hbar\omega_k^2 q_k g(\mathbf{k},\boldsymbol{\kappa})\}. \quad (3.21)$$

$$[\rho_{-k}\rho_{k'}, S]=\rho_{-k}[\rho_{k'}, S]+[\rho_{-k}, S]\rho_{k'}. \quad (3.22)$$

$$[\rho_{k'}, S]=i\sum_{\kappa,\kappa'} \{\delta_{\kappa\kappa'} c_{\kappa-k}^* c_{\kappa'-k'}-\delta_{\kappa'-k',\kappa-k} c_{\kappa'}^* c_\kappa^*\} \times\{f(\mathbf{k}',\boldsymbol{\kappa}')q_{k'}-ig(\mathbf{k}',\boldsymbol{\kappa}')p_{-k'}\}. \quad (3.23)$$

Diagonal terms are obtained if $\mathbf{k}'=\mathbf{k}$. If we keep only these, we may treat $[\rho_k, S]$ as a c-number. This is the procedure which Nakajima used, and is essentially equivalent to the Hartree approximation since there are evidently diagonal exchange terms in $\rho_{-k}[\rho_k, S]$ which are neglected. In this approximation, then,

$$\tfrac{1}{2}\sum_k M_k^2[\rho_{-k}\rho_k, S]=i\sum_{\kappa,k,s} M_k^2 \rho_{-k}\{n(\boldsymbol{\kappa}-\mathbf{k})-n(\boldsymbol{\kappa})\} \times\{f(\mathbf{k},\boldsymbol{\kappa})g_k - ig(\mathbf{k},\boldsymbol{\kappa})p_{-k}\}. \quad (3.24)$$

Elimination of linear terms in q_k and p_{-k} in the first line gives the following for f and g:

$$g(\mathbf{k},\boldsymbol{\kappa})=-\frac{\hbar f(\mathbf{k},\boldsymbol{\kappa})}{E_\kappa - E_{\kappa-k}}, \quad (3.25)$$

$$f(\mathbf{k},\boldsymbol{\kappa})=\frac{\hbar(E_\kappa - E_{\kappa-k})v_k}{(E_\kappa - E_{\kappa-k})^2-\hbar^2\omega_k^2}. \quad (3.26)$$

Principal parts are to be taken in the sums over the energy denominators. Elimination of linear terms in q_k from the second line then gives

$$\textstyle\sum_{k,s} M_k^2\{n(\boldsymbol{\kappa}-\mathbf{k})-n(\boldsymbol{\kappa})\}f(\mathbf{k},\boldsymbol{\kappa})=\hbar(v_k{}^i - v_k), \quad (3.27)$$

or, substituting for $f(\mathbf{k},\boldsymbol{\kappa})$,

$$v_k{}^i - v_k=-M_k^2 v_k\sum_{\kappa,s}\frac{(E_\kappa - E_{\kappa-k})(n(\boldsymbol{\kappa})-n(\boldsymbol{\kappa}-\mathbf{k}))}{(E_\kappa - E_{\kappa-k})^2-\hbar^2\omega_k^2} =M_k^2 v_k\sum_{\kappa,s}\frac{n(\boldsymbol{\kappa})-n(\boldsymbol{\kappa}-\mathbf{k})}{E_{\kappa-k}-E_\kappa+\hbar\omega_k}. \quad (3.28)$$

which can be solved readily to give v_k. Terms in p_{-k} vanish for a symmetrical distribution, such that $n(\boldsymbol{\kappa})=n(-\boldsymbol{\kappa})$.

Finally, elimination of diagonal terms in $q_k^* q_k$ gives the following equation for ω_k^2:

$$\Omega_k^2-\omega_k^2=-8v_{-k}{}^i v_k\sum_{\kappa,s}\frac{n(\boldsymbol{\kappa}-\mathbf{k})-n(\boldsymbol{\kappa})}{E_{\kappa-k}-E_\kappa+\hbar\omega_k}. \quad (3.29)$$

The $\boldsymbol{\kappa}$ on the left refers to the reduced vector in the first Brillouin zone. The sum 8 on the right is over all $\mathbf{k}$ which correspond to this same reduced wave vector, so that transitions due to Umklapp processes may be included. Contributions of electrons from different bands, if treated separately should also be added to the right-hand side. An alternative expression is

$$\Omega_k^2-\omega_k^2=8M_k^{-2}v_{-k}{}^i(v_k{}^i - v_k). \quad (3.30)$$

This is of the same form as Eq. (3.5). Equation (3.28) for the matrix element agrees with the 1937 result if $\hbar\omega_k$ in the denominator is neglected. Since $\hbar\omega_k \sim 10^{-4}E_F$, where E_F is the Fermi energy, the term will have practically no effect on the matrix element or frequency.

It should be noted that the Coulomb interaction between electrons still appears in the transformed Hamiltonian and we are still left with the problem of taking into account correlation effects on electron motion. The transformation does serve to introduce a set of oscillators which to second order are not coupled with the individual electrons. Scattering of electrons by oscillators is accounted for by interactions which conserve energy; i.e., those with zero energy denominators in the expansion. When summed by taking principal parts, these give a negligible contribution to the matrix element.

The exchange terms which Nakajima neglects are considered in Appendix B. Perhaps the most important effect of these terms is to add an exchange energy, W_κ, to the individual particle energy. It is known that such a term leads to an abnormally small density of states at the Fermi surface, and a correspondingly small low-temperature specific heat, which is contrary to observation.[10] This has been explained by the collective description in which the fields of the individual electrons are effectively screened by the other electrons. The exchange terms are drastically reduced, and one finds that the calculated specific heat is not far from the free-electron value neglecting exchange.[11] It is

[10] J. Bardeen, Phys. Rev. **50**, 1098 (1936).
[11] See P IV for a derivation of this result.

likely that inclusion of the unshielded exchange energy in the Nakajima calculation would lead to an incorrect electron-lattice interaction for long wavelengths. We shall show in Sec. 4 that the collective description is such as to introduce a shielded interaction into the calculation of the short-wavelength exchange terms, and no exchange effects at all appear for long wavelengths.

There remains the question of convergence of the expansion of the canonical transformation in a power series in v_k. The transformation is such as to introduce new individual particle wave functions which depend on the q_k's. The expansion coefficients are, neglecting $\hbar^2\omega_k^2$ in the denominator,

$$\hbar^{-1}f(\mathbf{k},\kappa)\cong\frac{v_kq_k}{E_{\kappa-k}-E_\kappa}. \tag{3.31}$$

One may expect that the expansion will converge rapidly if

$$\sum_k\frac{v_k^*v_kq_k^*q_k}{(E_{\kappa-k}-E_\kappa)^2}\ll 1. \tag{3.32}$$

The v_k are small and difficulty arises only from the terms for which the energy denominators are correspondingly small.

We shall now show that since principal parts are used, terms with small energy denominators contribute a negligible amount to the calculation of the matrix elements and frequency. Our analysis is similar to that of Frohlich[9] who suggested omitting from the transformation those $f(\mathbf{k},\kappa)$ which have small energy denominators, i.e., those for which

$$|E_{\kappa-k}-E_\kappa|<\Delta E. \tag{3.33}$$

When the same is done in our calculation, we find that we can choose ΔE large enough so that (3.32) is satisfied and small enough to have a negligible effect on the self-consistent field. To estimate the orders of magnitude involved, let us replace the matrix element by an average value, so that (3.32) becomes

$$|v_kq_k|_{\text{av}}^2\left\{\int_0^{E_F-\Delta E}\frac{N(E)dE}{(E-E_\kappa)^2}+\int_{E_\kappa+\Delta E}^{E_F}\frac{N(E)dE}{(E-E_\kappa)^2}\right\}\ll 1, \tag{3.34}$$

where $N(E)$ is the density of states in energy. This is roughly equivalent to

$$N(E_\kappa)|v_kq_k|_{\text{av}}^2\ll\Delta E. \tag{3.35}$$

If terms which satisfy (3.33) are omitted, what is the effect on the principal parts summation? The integral is of order:

$$\int_0^{E_F}\frac{N(E)dE}{E-E_\kappa}\sim N(E). \tag{3.36}$$

The omitted terms contribute roughly

$$\int_{E_\kappa-\Delta E}^{E_\kappa+\Delta E}\frac{N(E)dE}{E-E_\kappa}\cong 2\frac{dN}{dE}\Delta E. \tag{3.37}$$

The relative error is thus of order

$$\frac{2}{N(E)}\frac{dN}{dE}\Delta E\sim\frac{\Delta E}{E_F}. \tag{3.38}$$

The left-hand side of (3.35) is of order $\hbar\omega_{\max}\sim 10^{-4}E_F$ for most metals. Thus if one takes $\Delta E\sim 10^{-2}E_F$, one may satisfy (3.32) and at the same time not affect appreciably the calculation of the matrix element and vibrational frequency.

It is, however, just these omitted terms which are all important in the theory of superconductivity.[12] In fact, the criterion for the occurrence of superconductivity is roughly that

$$N(E)|v_kq_k|_{\text{av}}^2>\hbar\omega_{\max}. \tag{3.39}$$

The conclusion is that even though the electron-lattice interaction is so large as to give superconductivity, one may still use the expansion in powers of v_k for those terms which give almost the entire contribution to the matrix element and vibrational frequency. This is true because the virtual transitions which give difficulty are only a very small fraction of the total and do not contribute abnormally.

4. COLLECTIVE DESCRIPTION OF ELECTRON-ION INTERACTION

In this section we wish to treat the electron response to the ionic motion by describing the electrons in terms of the appropriate plasma variables. As pointed out by BP, such a description of the electrons is only appropriate for long wavelengths, corresponding to $k<k_c$, where k_c is a critical wave vector (or the order of k_0) which is determined by minimizing the total system energy as discussed in BP III. In this section we shall therefore only apply the collective description to the long-wavelength phonons, their interaction with the electrons, and the long-range electron-electron interactions (all corresponding to Fourier components with wave vectors less than k_c). We follow a procedure closely related to that used by one of us in dealing with the electron-electron interaction problem in the absence of ionic motion.[13]

Our basic aim in the collective description is the introduction of a new set of field variables which describe the independent plasma oscillations of the system as a whole. We may do this by first introducing a new field into our basic Hamiltonian, and then carrying out a series of canonical transformations which enable us to relate this field variable to the plasma oscillations we

[12] H. Frohlich, Phys. Rev. 79, 845 (1950); J. Bardeen, Revs. Modern Phys. 23, 261 (1951).

[13] D. Pines, Report to the Tenth Solvay Conference, Brussels (1954).

wish to describe. As a first step, we add a field-energy term to our initial Hamiltonian, (2.14) so that our Hamiltonian is now

$$H=H_1+\tfrac{1}{2}\sum_{|\mathbf{k}|<k_c} P_k^* P_k. \tag{4.1}$$

The P_k are here a quite arbitrary set of field variables, as yet undefined, which commute with all operators appearing in H_1. In order that the energy and number of degrees of freedom of our system remain unchanged by this field, we must further impose a set of subsidiary conditions on our combined system wave function:

$$P_k\Psi=0 \quad (|\mathbf{k}|<k_c). \tag{4.2}$$

We now wish to transform to a representation in which the P_k will describe independent plasma oscillations. In the absence of ionic motion, the ρ_k in the long-wavelength limit describe almost free collective oscillation, and the appropriate transformation would relate the P_k to the ρ_k. In our case, since the electrons also interact with the phonons, we would expect that it is only some linear combination of density fluctuations and phonon field variables which would carry out uncoupled collective plasma oscillation. Thus we are led to try a canonical transformation which relates the P_k to both ρ_k and the q_k. The first transformation we consider is generated by

$$S=\sum_{|\mathbf{k}|<k_c}(-iM_k\rho_{-k}+u_kq_{-k})Q_k, \tag{4.3}$$

where Q_k is the coordinate conjugate to the plasma field momentum P_k, and u_k is a real constant to be determined. After this transformation our subsidiary condition, (4.2) becomes

$$e^{-iS/\hbar}P_ke^{iS/\hbar}\Psi=[P_k-iM_k\rho_{-k}+u_kq_{-k}]\Psi=0. \tag{4.4}$$

Our Hamiltonian (4.1) is transformed to

$$\begin{aligned}H=&\sum_\kappa E_\kappa c_\kappa^* c_\kappa+\tfrac{1}{2}\sum_{|\mathbf{k}|<k_c}\{p_k^*p_k+(\Omega_k^2-\omega_k^2)q_k^*q_k\}\\&+\tfrac{1}{2}\sum_{|\mathbf{k}|<k_c}\{P_k^*P_k+(\omega_p^2+u_k^2)Q_k^*Q_k\}\\&+\sum_{|\mathbf{k}|<k_c}\{v_k{}^i-iM_ku_k\}q_k\rho_k+\sum_{|\mathbf{k}|<k_c}u_k\rho_k^*Q_k\\&-\sum_{|\mathbf{k}|<k_c,\kappa}M_\kappa\frac{\hbar\mathbf{k}}{m}\cdot(\boldsymbol{\kappa}-\tfrac{1}{2}\mathbf{k})c_\kappa^*c_{\kappa-k}Q_k\\&+\tfrac{1}{2}\sum_{|\mathbf{k}|>k_c}\{p_k^*p_k+\Omega_k^2q_k^*q_k\}+\sum_{|\mathbf{k}|>k_c}v_k{}^iq_k\rho_{-k}\\&+\tfrac{1}{2}\sum_{|\mathbf{k}|>k_c}M_k^2\rho_{-k}\rho_k.\end{aligned} \tag{4.5}$$

In obtaining this expression we have made use of the new subsidiary condition, (4.4), and we have applied the "random phase" approximation of BP.[14] We have also applied the effective mass approximation, $E(\boldsymbol{\kappa})=\hbar^2\boldsymbol{\kappa}^2/2m$.[15]

The first three terms of our new Hamiltonian, (4.5), describe the one-electron energy levels, the phonon field, and the plasma field. The next three describe the electron-phonon interaction, plasma-phonon interaction, and plasma-electron interaction, while the remaining terms describe the short-wavelength phonons, their interaction with the electrons, and the short-range electron-electron interaction. We see that we have here redescribed the long-range electron-electron interactions in terms of the plasma oscillations, which however are not yet "isolated," in the sense that there remains a plasma-electron and plasma-phonon interaction, and the plasma field variable still appears in the subsidiary condition. The effect of the electron-electron interactions on the effective interaction matrix element and on the sound wave frequency is contained in the u_k, which as we see measure the coupling between the plasma waves and the phonons. The u_k are simply related to the $v_k{}^\rho$ we introduced in the preceding section to describe the effect of electron shielding on the v_k; we have

$$v_k{}^\rho=-iM_ku_k. \tag{4.6}$$

Our remaining problem is thus the determination of u_k and the new sound wave frequencies.

We may actually do this without explicitly solving for the effect of the electron-plasma interaction and plasma-phonon interaction on the problem at hand. For, as is the case in the absence of ionic motion, the coupling between the electrons and the plasma oscillations is weak. Furthermore, because of the great disparity between the plasma frequencies and phonon frequencies (a disparity of order m/M), the coupling between the plasma waves and the phonons is likewise weak. Both of these interactions may easily be taken into account by appropriate canonical transformations; however because of the weakness of the coupling, such transformations will leave the relevant terms in our Hamiltonian essentially unaltered, so we do not need to carry them out explicitly here.

How, then, do we choose the u_k? It is clear from the discussion of the preceding section that we must choose the u_k in such a way that our treatment is self-consistent. The requirement of self-consistency appears in the following guise. In our problem the electrons and phonons are coupled not only in the Hamiltonian (4.8) but in the subsidiary condition (4.9) as well. The requirement of self-consistency is then just the requirement that the coupling via the subsidiary condition be completely equivalent to the coupling via the Hamil-

[14] If the phonon terms in (4.5) are suppressed (and $u_k=0$), it may be seen that the form of (4.5) is identical with the Hamiltonian used in BP III as the starting point for the collective description.

[15] The use of the effective mass approximation for this problem is discussed by J. Hubbard, Proc. Phys. Soc. (London) **A67**, 1058 (1954); P. Wolff, Phys. Rev. **95**, 56 (1954). E. N. Adams II [Phys. Rev. **98**, 947 (1955)] has pointed out that interband transitions may have an important effect. N. F. Mott [Proceedings of the Tenth Solvay Conference, Brussels (1954)] has given reasons for expecting that the free electron mass rather than an effectual mass would appear in many cases.

tonian. This guarantees that we introduced the correct admixture of phonon coordinates (as determined by the electron-phonon interaction) in the plasma oscillation amplitude. We obtain an explicit solution for the u_k and the sound-wave frequencies ω_k by carrying out a canonical transformation which eliminates to a given order of accuracy the electron-phonon interaction terms in the Hamiltonian, (4.5). Our self-consistency requirement is then that to this same order of accuracy there be no coupling of the sound waves to the electrons in the transformed subsidiary condition; this will be the case if the phonon variable no longer appears in the transformed subsidiary condition in this order.

Our desired canonical transformation is just that generated by (3.19), since we have identified v_k with $v_k{}^i - iM_k\rho_{-k}$. As before $f(\mathbf{k},\boldsymbol{\kappa})$ and $g(\mathbf{k},\boldsymbol{\kappa})$ are defined through (3.25) and (3.26), and as discussed in the preceding section a perturbation theory expansion in powers of v_k is valid. Our transformed Hamiltonian is[16]

$$\begin{aligned}H=&\sum_{\kappa} E_\kappa c_\kappa{}^* c_\kappa+\tfrac{1}{2}\sum_{|\mathbf{k}|<k_c}\{p_k{}^* p_k+\omega_k{}^2 q_k{}^* q_k\}\\&+\tfrac{1}{2}\sum_{|\mathbf{k}|<k_c}\{P_k{}^* P_k+(\omega_p{}^2+u_k{}^2)Q_k{}^* Q_k\}\\&-\sum_{|\mathbf{k}|<k_c,\kappa} M_\kappa\frac{\hbar\mathbf{k}}{m}\cdot(\boldsymbol{\kappa}-\tfrac{1}{2}\mathbf{k})c_\kappa{}^* c_{\kappa-k}Q_k\\&+\tfrac{1}{2}\sum_{|\mathbf{k}|>k_c}(p_k{}^* p_k+\Omega_k{}^2 q_k{}^* q_k)+\sum_{|\mathbf{k}|>k_c} v_k{}^i q_k\rho_{-k}\\&+\tfrac{1}{2}\sum_{|\mathbf{k}|>k_c} M_k{}^2\rho_k\rho_{-k}\\&-\tfrac{1}{2}\sum_{|\mathbf{k}|<k_c,\kappa} v_{-k}\rho_k c_\kappa{}^* c_{\kappa-k}g(\mathbf{k},\boldsymbol{\kappa}).\end{aligned}\tag{4.7}$$

The dispersion relation for our "uncoupled" phonons is

$$\omega_k{}^2=\Omega_k{}^2-u_k{}^2+v_k{}^2\sum_\kappa\frac{n(\boldsymbol{\kappa}-\mathbf{k})-n(\boldsymbol{\kappa})}{\hbar\omega_k+E_{\kappa-k}-E_\kappa}.\tag{4.8a}$$

We further find, using (3.23) that our self-consistency requirement that q_k no longer appear in the subsidiary condition yields:

$$u_k=-iM_k v_k\sum_\kappa\frac{n(\boldsymbol{\kappa}-\mathbf{k})-n(\boldsymbol{\kappa})}{\hbar\omega_k+E(\boldsymbol{\kappa}-\mathbf{k})-E(\boldsymbol{\kappa})}.\tag{4.8b}$$

On applying (4.6) we see that this choice of u_k (or v_k) is in complete agreement with the result of Nakajima (3.28). Furthermore, on combining (4.8a) with (4.8b) we obtain the by now familiar dispersion relation (3.5).

A transformation similar to that of Nakajima may be used to eliminate the terms linear in q_k for $|\mathbf{k}|>k_c$. The result of both transformations is to replace the electron-lattice interaction by an interaction between electrons:

$$\begin{aligned}H_2=&-\tfrac{1}{2}\sum_{\kappa,|\mathbf{k}|<k_c} v_{-k}\rho_k c_\kappa{}^* c_{\kappa-k}g(\mathbf{k},\boldsymbol{\kappa})\\&-\tfrac{1}{2}\sum_{\kappa,|\mathbf{k}|>k_c} v_{-k}{}^i\rho_k c_\kappa{}^* c_{\kappa-k}g(\mathbf{k},\boldsymbol{\kappa}).\end{aligned}\tag{4.9}$$

[16] We have omitted the weak plasma-phonon interaction term. Note that the unshielded interaction, $v_{-k}{}^c$, appears from the Nakajima transformation. Since $v_{-k}{}^c$ becomes infinite while v_{-k} approaches zero as $\mathbf{k}\to 0$, the difference is particularly important when $\mathbf{k}$ is small.

The interaction term, H_2, is similar to one derived by Frohlich[9] without explicit introduction of Coulomb interactions. His interaction constant is to be identified with the shielded interaction, v_k. As he points out, the usual second-order perturbation theory expression for the electron-lattice interaction energy is the sum of the diagonal component of H_2 (which Frohlich calls E_2) and the change in zero-point energy of the oscillators. It is the true frequency, ω_k, rather than Ω_k, which enters into these expressions.

There are several other important differences between our treatment and that of Nakajima. In (4.7) the long-range Coulomb interactions no longer appear, but have been redescribed by the high-frequency plasma oscillations. There remains the weak plasma-electron interaction bu. this may be easily transformed away, as is done in BP III, and it may then be seen that only effective residual electron-electron interaction is that described by

$$H_{s.r.}=\tfrac{1}{2}\sum_{|\mathbf{k}|>k_c} M_k{}^2\rho_k\rho_{-k}.$$

Thus the long-range electron-electron correlations have been described in terms of the plasma oscillations, and are seen explicitly not to affect the electron response to the long wavelength phonons. We further note that no exchange effects are associated with the long-range part of the electron-electron interactions, so that the use of the Hartree approximation in determining v_k for $|\mathbf{k}|<k_c$ is completely justified.

There remains the complication introduced by the transformed subsidiary condition, which on applying (4.8b) agrees with that for the electron-plasma system in the absence of ionic motion:

$$(P_k-iM_k\rho_{-k})\Psi=0.\tag{4.10}$$

This complication has been discussed in BP III. The same transformation which eliminates the plasma-electron interaction in the Hamiltonian (4.7) also eliminates the plasma momentum from the subsidiary condition (4.10). There remains then a subsidiary condition acting on the electrons alone. As shown in BP III this subsidiary condition will be automatically satisfied for the lowest state of the system. It will influence the excited states in such a way as to reduce the effective number of electronic degrees of freedom slightly, but because the number of plasma degrees of freedom is relatively small (of order 5%–10%) this reduction should be correspondingly small.

We still have in our Hamiltonian (4.7) the short-wavelength terms corresponding to $|\mathbf{k}|>k_c$; the short-wavelength phonons, their interaction with the electrons, and the short-range part of the electron-electron interaction. As mentioned earlier, these terms may be

treated by the method of Nakajima in order to determine the effective electron-lattice interaction and sound wave frequencies. As discussed in Appendix B, for this wavelength region exchange effects are of importance. However, they are in principle calculable and should not lead to anomalous results, since the effective electron interaction is screened. Because of the mathematical complexity of the system of coupled equations obtained in Appendix B, we have not been able to obtain an explicit solution for the short wavelength v_k and ω_k when exchange effects are included.

The short-range electron-electron interaction occurring in (4.7) will not have an appreciable effect on the electronic wave functions. For as is shown in P IV, the range of this interaction is of the order of the interparticle spacing, and it may consequently in fact be treated as a relatively small perturbation on the electronic motion, and thus on the electronic wave functions.

5. CONCLUSIONS

We are led to the following physical picture of the coupled motion of ions and electrons in a metal. As the ions move, the electrons tend to follow their motion so as to screen out the ion field in a distance comparable with the interparticle spacing. Thus the fluctuations in potential caused by the motion of the ions are greatly reduced. The effective matrix element for electron-lattice interaction is determined by a screened field which can be calculated rather accurately for long wavelengths by elementary considerations. The positions of the electrons are correlated so that the field of a given electron is also screened out in a distance of the order of the interparticle spacing. Long-range correlation effects, according to the collective description, give rise to plasma waves and to coupled ion-electron waves which correspond to longitudinal sound waves. The remaining short-range interactions can be taken into account in a satisfactory way by the individual particle model.

In the collective description, extra coordinates are introduced to describe the plasma oscillations, and introduction of these requires that the system wave function satisfy supplementary conditions. A series of canonical transformations is made in order to isolate the plasma oscillations and the longitudinal sound waves. The effective matrix element for electron-phonon interaction is thereby determined and found to be identical for long wavelengths with that found earlier by self-consistent field methods which neglect effects of exchange and correlation. There remain in the transformed Hamiltonian terms which describe individual particles which interact with each other with a screened Coulomb force and are coupled with the phonon field in the usual way. The screened interaction is sufficiently weak so that it can be treated by perturbation methods and does not have a large effect on the particle wave functions. In this way we see that the usual empirical individual-particle model is justified.

An expression is derived for phonon frequencies of long wavelength and is found to be identical with that derived by Toya and Nakajima by self-consistent field methods. An explicit calculation for monovalent metals, given in an appendix, yields, when applied to Na, good agreement with values calculated from observed elastic constants. Our calculation makes use of matrix elements derived in A by the self-consistent field method. The elastic constants may also be obtained by a direct calculation of the energy of a distorted crystal, as has been done by Fuchs. The close agreement both with the observed values and with the direct calculation makes one confident that the general method for calculating matrix elements and vibrational frequencies is correct.

The calculation of the matrix elements and vibrational frequencies for short wavelengths is less reliable, because we have not been able to take into account the influence of exchange on these quantities. Perhaps the best procedure in calculations of conductivity at room temperatures, where short wavelength phonons and Umklapp processes are important, is to introduce an empirical scattering factor, as has recently been suggested by Ziman.[17]

In the formulation of the superconductivity problem, it is probably a reasonably good approximation to use an empirical interaction constant, v_k, and to omit explicit introduction of Coulomb interaction terms, as has been done by Frohlich and by Bardeen. The fact that v_k^2 rather than v_k appears in the interaction Hamiltonian, H_2, in (4.9) for $|\mathbf{k}|>k_c$ raises some doubt about this procedure, but the difference is not large for these short wavelengths.

Except for terms which correspond to transitions which nearly satisfy the conservation of energy,

$$|E_{\kappa'}-E_{\kappa}\pm\hbar\omega_k|<\Delta E, \tag{5.1}$$

the electron-lattice interaction can be eliminated by a canonical transformation such that in the final Hamiltonian the lattice oscillators are not coupled with the electrons. It is the remaining interaction terms which do satisfy (5.1) that are responsible for scattering of electrons and also presumably account for superconductivity. The value of ΔE always can be chosen sufficiently small so that these terms have a negligible effect on the electron-lattice matrix element and on the vibrational frequency. On the other hand, they cannot be treated by perturbation theory, and so can have a pronounced effect on the electron wave functions.

In the theory of superconductivity, then, one need only consider virtual transitions which satisfy an expression of the form (5.1). This approach was followed by one of the authors,[12] with ΔE chosen to be of the order of the electron-lattice interaction energy resulting

[17] J. M. Ziman, Proc. Roy Soc. (London) **A226**, 432 (1954).

from these virtual transitions. This gives [see Eq. (3.35)]

$$\Delta E \sim N(E)\langle|v_k q_k|^2\rangle_{\text{Av}}, \tag{5.2}$$

which is an order of magnitude or so larger than kT_c (T_c = transition temperature) for most superconductors. If one assumes that these interactions contribute to the superconducting state, but not to the normal state, one would have far too large an energy difference between the two phases. About the correct order of magnitude for this energy difference is obtained if we arbitrarily take $\Delta E \sim kT_c$; that is if electrons with energies within $\sim kT_c$ of the Fermi surface have their energies lowered by $\sim kT_c$. Undoubtedly the virtual transitions determined by (5.2) contribute to both the normal and superconducting states, and only a very small fraction of the interaction energy is involved in the change of state. It is evident that better pictures of both the normal and superconducting phases are required. The equations we have presented here should provide a good basis for development of an adequate theory.

APPENDIX A: CALCULATION OF ω_k FOR A MONOVALENT METAL

We give here an explicit calculation of the sound-wave frequencies for the long wavelength limit, $\mathbf{k}\to 0$, by making use of the matrix elements of the electron-lattice interaction in Eq. (3.12). The Nakajima expression, Eq. (3.30), gives equivalent results in this limit. The matrix elements used are those derived originally for a calculation of the conductivity of monovalent metals.

Determining the frequencies for long wavelengths is equivalent to calculating the elastic constants, which are usually obtained by a direct calculation of the energy of a distorted crystal. Compressibilities of monovalent metals have been determined by a Wigner-Seitz calculation of the energy as a function of volume. Fuchs[18] has shown that the shear constants of the alkali metals are given quite closely by the Coulomb interactions of the ions in a uniform negative sea. Although important for the noble metals, repulsion between the closed shells is almost negligible for the alkalis.

It is most convenient to determine the sum of the squares of the frequencies for the three different directions of polarization, since thus sum is independent of the direction of propagation. Kohn[19] has shown that the sum for the Coulomb interactions of the ions alone is just the square of the plasma frequency for a gas of the same density:

$$\sum_{\sigma=1,2,3} \Omega_{k\sigma}^2 = \frac{4\pi N e^2}{M}. \tag{A1}$$

Only the longitudinal component is affected by the electron-lattice interaction, so that, from Eq. (3.12),

$$\sum_{\sigma} \omega_{k\sigma}^2 = \sum_{\sigma} \Omega_{k\sigma}^2 - \frac{k^2}{4\pi e^2}|v_k^i|^2\left(1-\frac{k^2 E_F}{6\pi N e^2}\right), \tag{A2}$$

where

$$E_F = \hbar^2 k_0^2/2m \tag{A3}$$

is the Fermi energy. In terms of the elastic constants

$$\textstyle\sum_\sigma \omega_{k\sigma}^2 = NMk^2(c_{11}+2c_{44}). \tag{A4}$$

The expression for the matrix element v_k^i is taken from reference 1,

$$v_k^i = -\frac{ik}{(NM)^{\frac{1}{2}}}\left\{\frac{4\pi N e^2}{k^2}+\gamma(V_0(r_s)-E_0)\right\} \times\left\{\frac{3(\sin kr_s - kr_s\cos kr_s)}{(kr_s)^3}\right\}, \tag{A5}$$

where r_s is the radius of a sphere of atomic volume, $V_0(r_s)\cong 0$ is the potential at the boundary of the s-sphere, E_0 is the energy of the lowest state, and $\gamma = |\psi_0(r_s)|^2/\langle|\psi_0(r)|^2\rangle_{\text{Av}}$. Both V_0 and E_0 refer to a potential in which the field of the ions is compensated by a uniform negative charge. Wigner and Seitz calculated the energy, $\mathcal{E}_0$, for a cell in which the field is that of the bare ion at the center. If $v_0(r_s)$ is the corresponding potential, there is the approximate relation

$$V_0(r_s)-E_0 = v_0(r_s)-\mathcal{E}_0-0.2e^2/r_s. \tag{A6}$$

There is also the following relation, originally derived by Frohlich:

$$\frac{r_s}{3}\frac{d\mathcal{E}_0}{dr_s} = \gamma(v_0(r_s)-\mathcal{E}_0). \tag{A7}$$

An expansion of (A5) in a power series in k is

$$\begin{aligned} v_k^i &= -i\frac{4\pi e^2}{k}\left(\frac{N}{M}\right)^{\frac{1}{2}}\left\{1+\frac{k^2\gamma}{4\pi N e^2}[V_0(r_s)-E_0] - \frac{1}{10}k^2 r_s^2+\cdots\right\} \\ &= -i\frac{4\pi e^2}{k}\left(\frac{N}{M}\right)^{\frac{1}{2}}\left[1+\frac{k^2}{4\pi N e^2} \times\left[\gamma[V_0(r_s)-E_0]-\frac{3}{10}\frac{e^2}{r_s}\right]+\cdots\right\}, \end{aligned} \tag{A8}$$

where we have used $N = 3/4\pi r_s^3$. Substitution of (A8) into (A2) gives after some reduction

$$\sum_\sigma \omega_{k\sigma}^2 = \frac{k^2}{M}\left\{\frac{2}{3}E_F+\frac{6e^2}{10r_s}-2\gamma[V_0(r_s)-E_0]+\cdots\right\},$$

[18] K. Fuchs, Proc. Roy. Soc. (London) **153**, 662 (1936); **157**, 444 (1936).

[19] W. Kohn (private communication). We should like to thank Professor Kohn for communicating his results to us in advance of publication.

so that, from (A4),

$$c_{11}+2c_{44}=N\left\{\frac{2}{3}E_F+\frac{6e^2}{10r_s}-2\gamma[V_0(r_s)-E_0]+\cdots\right\}.$$

Inserting appropriate values for sodium, with the free-electron value for E_F and $\gamma=1$, we find

$$2.6\times10^{22}\{3.3+6.6-0.6\}\times10^{-12}=2.4\times10^{11}\ \text{ergs/cm}^3,$$

while

$$c_{11}+2c_{44}=\{0.95+2(0.59)\}\times10^{11}=2.13\times10^{11}\ \text{ergs/cm}^3.$$

The difference between calculated and observed values is only about 10%. This agreement gives one confidence in the method of calculation and in the values of the matrix elements for electron-lattice scattering for small k.

APPENDIX B: EXCHANGE TERMS FROM COULOMB INTERACTION

There are a number of exchange terms in the commutator of the Coulomb interaction:

$$\tfrac{1}{2}\sum M_{k'}{}^2[\rho_{-k'}\rho_{k'},S]=\tfrac{1}{2}\sum M_{k'}{}^2 \times\{\rho_{-k'}[\rho_{k'},S]+[\rho_{-k'},S]\rho_{k'}\}, \quad \text{(B1)}$$

which were omitted in Nakajima's treatment. Exchange terms appear only for electrons of parallel spins. Expansion of the commutators for these gives

$$\frac{i}{2}\sum M_{k'}{}^2\{c_{\kappa'}{}^*c_{\kappa'-k'}(c_{\kappa-k'}{}^*c_{\kappa-k}-c_\kappa{}^*c_{\kappa-k+k'}) +(c_{\kappa-k'}{}^*c_{\kappa-k}-c_\kappa{}^*c_{\kappa-k+k'})c_{\kappa'}{}^*c_{\kappa'-k'}\} \times\{f(\mathbf{k},\kappa)q_k-ig(\mathbf{k},\kappa)p_{-k}\}. \quad \text{(B2)}$$

In the second line we have changed the sign of $\mathbf{k}'$.

Our problem is to find diagonal parts of the coefficients of $c_\kappa{}^*c_{\kappa-k}q_k$ and corresponding terms. In addition to those for which $\mathbf{k}'=\mathbf{k}$, κ' arbitrary, used in obtaining (3.24), there are several others: (a) $\kappa'=\kappa$, (b) $\kappa'=\kappa-\mathbf{k}$, (c) $\kappa=\kappa'-\mathbf{k}'$, (d) $\kappa'=\kappa-\mathbf{k}+\mathbf{k}'$. The sum of these may be expressed in the form

$$\begin{aligned}\tfrac{1}{2}\sum M_{k'}{}^2[\rho_{-k'}\rho_{k'},S]=i\sum_{\kappa,k}\{&\sum_{\kappa',s}M_k{}^2 \times\{n(\kappa'-\mathbf{k})-n(\kappa')\}f(\mathbf{k},\kappa')\\ &+\sum_{k'}M_{k'}{}^2\{n(\kappa-\mathbf{k}-\mathbf{k}')-n(\kappa-\mathbf{k}')\}f(\mathbf{k},\kappa)\\ &+\sum_{\kappa'}M_{\kappa'-k}{}^2\{n(\kappa')-n(\kappa'-\mathbf{k})\}f(\mathbf{k},\kappa')\}c_\kappa{}^*c_{\kappa-k}q_k\\ &+\text{corresponding terms in } g(\mathbf{k},\kappa). \quad \text{(B3)}\end{aligned}$$

The first sum in the curly brackets is over both spin states, the other two only over the spin which is parallel to that of $c_\kappa{}^*c_{\kappa-k}$.

When exchange terms are included, the effects of the other electrons on a given electron can no longer be expressed in terms of a potential, and it is no longer advantageous to introduce v_k. In order to eliminate linear terms in q_k and p_{-k} in the transformed Hamiltonian, $f(\mathbf{k},\kappa)$ and $g(\mathbf{k},\kappa)$ must satisfy the following equations:

$$\begin{aligned}&\textstyle\sum_{\kappa'}(2M_k{}^2-M_{\kappa-\kappa'}{}^2)[n(\kappa'-\mathbf{k})-n(\kappa')]f(\mathbf{k},\kappa') -(W_{\kappa-k}-W_\kappa)f(\mathbf{k},\kappa)+(E_\kappa-E_{\kappa-k})f(\mathbf{k},\kappa)\\ &\qquad+\hbar\omega_k{}^2g(\mathbf{k},\kappa)+v_k{}^i=0,\\ &\textstyle\sum_{\kappa'}(2M_k{}^2-M_{\kappa-\kappa'}{}^2)[n(\kappa'-\mathbf{k})-n(\kappa')]g(\mathbf{k},\kappa') -(W_{\kappa-k}-W_\kappa)g(\mathbf{k},\kappa)+(E_\kappa-E_{\kappa-k})g(\mathbf{k},\kappa)\\ &\qquad+\hbar f(\mathbf{k},\kappa)=0, \quad \text{(B4)}\end{aligned}$$

where W_κ is the exchange energy of an electron in the state κ:

$$W_\kappa=-\textstyle\sum_{k'}M_{k'}{}^2n(\kappa-\mathbf{k}'). \quad \text{(B5)}$$

The factor two multiplying $M_k{}^2$ takes account of the sum over spins. The equations are such that for a symmetrical distribution of electrons in κ space, $n(\kappa)=n(-\kappa)$, the solutions satisfy the relations:

$$f(\mathbf{k},\mathbf{k}-\kappa)=-f(\mathbf{k},\kappa);\quad g(\mathbf{k},\mathbf{k}-\kappa)=g(\mathbf{k},\kappa). \quad \text{(B6)}$$

The direct term in the equation for $g(\mathbf{k},\kappa)$ then vanishes:

$$\textstyle\sum_{\kappa'}M_{k'}{}^2[n(\kappa'-\mathbf{k})-n(\kappa')]g(\mathbf{k},\kappa')=0. \quad \text{(B7)}$$

When the exchange terms are included, the equations cannot be solved algebraically.

There are two types of exchange terms. One simply adds the exchange energy W_κ to the individual particle energy E_κ. The other adds sums over $f(\mathbf{k},\kappa')$ and $g(\mathbf{k},\kappa')$ to the equations for f and g, and it is these which make the solution difficult. An estimate of the magnitude of the latter terms can be obtained by taking an average over κ. We have

$$\textstyle\sum_\kappa M_{\kappa-\kappa'}{}^2n(\kappa)=-W_{\kappa'}. \quad \text{(B8)}$$

An average of $W_{\kappa'}$ is the exchange energy, $-0.92e^2/r_s$ for a monovalent metal. This is to be compared with

$$2M_k{}^2\textstyle\sum n(\kappa)=n_0M_k{}^2=4\pi n_0e^2/k^2=(3e^2/r_s)(kr_s)^{-2}, \quad \text{(B9)}$$

where $n_0=3/4\pi r_s{}^3$ is the concentration of electrons. These exchange terms may be neglected when $kr_s\ll1$, i.e., for long wavelengths. However, as discussed in the text, the exchange energies of the individual electrons, W_κ, would make marked changes at long wavelengths. As the collective description shows, these should not be included.

Intrinsic Optical Absorption in Single-Crystal Germanium and Silicon at 77°K and 300°K*

W. C. Dash and R. Newman
General Electric Research Laboratory, Schenectady, New York

(Received May 2, 1955)

The intrinsic absorption spectra of high-purity single-crystal germanium and silicon have been measured at 77°K and 300°K. The spectral regions studied encompassed a range of absorption coefficient from 0.1 cm^{-1} to 10^4 cm^{-1} for each material. The germanium data may be interpreted as indicating a threshold for direct transitions at 0.81 ev at 300°K and at 0.88 ev at 77°K. The threshold for indirect transitions was placed at 0.62 ev and 0.72 ev for 300°K and 77°K, respectively. For silicon the data were not as readily interpreted However, there is an indication that the threshold for direct transitions should be placed at about 2.5 ev and the threshold for indirect transitions at 1.06 ev and 1.16 ev at 300°K and 77°K, respectively.

INTRODUCTION

RECENT advances in our knowledge of the band structure of germanium and silicon[1–3] have revived interest in the problem of the intrinsic optical absorption process in these materials.[4–6] A theory has been developed by Hall, Bardeen, and Blatt which attempts to relate the optical absorption coefficient to the fundamental parameters of the band structure (e.g., shape and separation of the band surfaces).[7] This theory envisions two processes to account for the absorption. In the first process, absorption of the photon takes place with k-vector being conserved within the electron system itself (i.e., vertical transitions). For this "direct" process, the low-energy threshold for absorption would occur at the photon energy corresponding to the minimum energy separation, measured vertically on a reduced zone plot, between valence and conduction band states. This minimum separation is probably located at the origin of the Brillouin zone for both germanium and silicon.

For the second process, the indirect process, k-vector is not conserved within the electron system, but rather conservation is obtained by the emission or absorption of phonons. The process is pictured as a diagonal transition on a reduced zone plot. The threshold energy for the indirect transitions would correspond to the minimum energy separation, without restriction, between valence and conduction bands, less the energy of a phonon having the required k vector.

If the two thresholds are well separated, one might expect to see a separation of the absorption spectrum into two regions, one in which the absorption process is dominated by direct transitions, the other for which indirect transitions would be dominant.

Previous measurements of the absorption coefficient in single crystal silicon and germanium were restricted to absorption coefficients less than about 10^3 cm^{-1} because of problems of sample thickness.[4–6] By using evaporated films, it was possible to study higher absorption coefficients.[5] Some early optical investigations indicated to us that the values obtained from such films were not characteristic of single-crystal material. In view of the theoretical interest in the subject, it seemed worthwhile to develop methods that would allow for as complete a study of absorption coefficients in single crystal materials as could be made. Accordingly the study described below was undertaken.

EXPERIMENTAL

Samples of high-purity germanium and silicon with resistivities in excess of 20 ohm-cm were cut to an appropriate thickness with a diamond saw. The faces were ground with successively finer grades of carborundum and then polished with aluminum oxide polishing powder on a bee's wax lap. For samples whose thickness was 100 microns or less, special pains were necessary to obtain samples that were of uniform thickness and free from pinholes. For these samples one face was first ground and polished so that no scratch marks were visible under a low-powder microscope. This face was then pressed in contact with a glass plate on which a resin had been spread. Thus mounted, the samples were ground and polished down to the desired thickness. Alkanex, a thermosetting resin, was used for samples 0.5μ to 2μ thick. Glycol phthalate, a thermoplastic resin, was used for samples 2μ or thicker. The latter had the advantage of permitting the sample to be freed from the backing for the measurement.

* A preliminary report of this work was presented at the 1955 New York Meeting of the American Physical Society [Phys. Rev. **98**, 1536(A) (1955)].

[1] F. Herman, Phys. Rev. **88**, 1210 (1952); F. Herman and Callaway, Phys. Rev. **89**, 518 (1953); F. Herman, Phys, Rev. **93**, 1214 (1954); F. Herman, Phys. Rev. **95**, 847 (1954); F. Herman, Physica **20**, 801 (1954).

[2] Dresselhaus, Kip, and Kittel, Phys. Rev. **92**, 827 (1953); Dresselhaus, Kip, and Kittel, Phys. Rev. **95**, 568 (1954); C. Kittel, Physica **20**, 829 (1954).

[3] Lax, Zeiger, Dexter, and Rosenblum, Phys. Rev. **93**, 1418 (1954); Dexter, Zeiger, and Lax, Phys. Rev. **95** 557 (1954); Dexter, Lax, Kip, and Dresselhaus, Phys. Rev. **96**, 222 (1954); R. N. Dexter and B. Lax, Phys. Rev. **96**, 223 (1954); Lax, Zeiger, and Dexter, Physica **20**, 818 (1954).

[4] H. Y. Fan and M. Becker, in *Semiconducting Materials* (Butterworth Publications, London, 1951), p. 132; Fan, Sheppard, and Spitzer, Atlantic City Photoconductivity Conference, November, 1954 (to be published).

[5] W. H. Brattain and H. B. Briggs, Phys. Rev. **75**, 1705 (1949).

[6] H. B. Briggs, J. Opt. Soc. Am. **42**, 686 (1952).

[7] Hall, Bardeen, and Blatt, Phys. Rev. **95**, 559 (1954); also Atlantic City Photoconductivity Conference (unpublished).

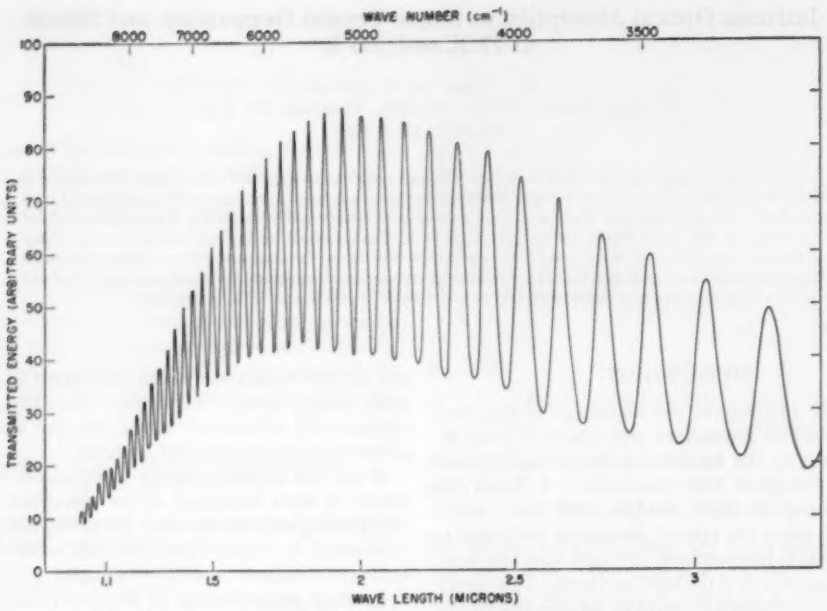


FIG. 1. Interference fringe pattern for a silicon sample 9.2μ thick and about 5×5 mm in area.

Thickness was measured, variously, by direct measurement with a dial gauge (0.1 mil/division), weighing, or measurement of the interference fringe spectrum. The last method gave the most reliable results for thin samples and also some measure of the nonuniformity of the samples. If, for example, a nonuniformity in thickness can be described by saying that the sample consists of two sections of equal area, one of which is thicker than the other by an amount δ, then it is possible to calculate very simply the quantity δ from the intensity pattern of the interference fringes.[8] For one sample of silicon of mean thickness equal to 9.2μ and an area of about 5×5 mm, a δ of $\sim$0.07μ was calculated from the fringe pattern which is shown in Fig. 1. In practice, nonuniformities in thickness of about 5 percent are unimportant in the absorption constant measurements. Samples used for these measurements were well within this tolerance.

The optical measurements were made on a Perkin-Elmer 112 double-pass spectrometer using a CaF_2 prism. Room temperature measurements were made with samples mounted at either the entrance or exit port. No difference was observed between the two measurements. The liquid nitrogen measurements reported were made with the samples mounted in a transparent dewar flask and immersed in the refrigerant. It was found that this method gave as reliable data as more elegant methods and had the further advantage of insuring that the samples were at temperature, a far from trivial point in the case of the very thin samples.

[8] M. Born, *Optik* (Verlag Julius Springer, Berlin, 1933), p. 122.

Depending on the spectral region being studied, a thermocouple, a PbS cell, or a photomultiplier was used as a detector. Where overlapping data points were available, the agreement between different measurements was good.

Absorption measurements were made by the usual sample-in, sample-out technique correcting for the reflectance of the samples. Reliable absorption coefficients could usually be obtained with the measured transmission in the range from about 0.3 to 10^{-3}. For very low transmissions it was necessary to check for stray light signals and nonlinearity in the over-all detection system. In some instances, absorption coefficients, which subsequently proved to be reliable, were obtained from transmissions as low as 10^{-6}. Because our measurements were made in the low-transmission range, the corrections for the effects of sample backing described by Brattain and Briggs[5] were unimportant. Data taken on a given sample with and without its backing gave identical absorption coefficients.

An important aspect of the work was concerned with determining to what extent the thinnest samples had the properties of bulk single-crystal material. To this end, two experiments were performed. The first was to obtain an x-ray diffraction picture and observe whether the thinnest material was still single crystal and to what extent it was disordered. Back-reflection Laue photographs of samples $\sim$1μ thick indicated single-crystal patterns which, to the resolution of the x-ray camera, were indistinguishable from the patterns for bulk single crystals.

The second experiment was perhaps a more sensitive test of the crystalline perfection of the thin samples. It consisted of a measurement of the Hall effect in some of these samples. A germanium sample of 2μ in thickness was found to be p-type with a hole mobility at room temperature of about 600 cm²/volt sec. After an annealing treatment, the mobility increased to about 2500 cm²/volt sec. The hole mobility in the parent material, which was p-type, was also about 2500 cm²/volt sec. A silicon sample 2μ thick, which had been cut from a p-type crystal, was found to be n-type with a mobility of about 500 cm²/volt sec at room temperature. The electron mobility in good silicon is of the order 1500 cm²/volt sec. The effect of annealing was not measured on this sample. No difference was observed optically between other annealed and unannealed samples at either 300°K or 77°K.

Summarizing the results given above, we conclude that the thin samples are single crystal, although probably not of the perfection of the bulk material. We have, unfortunately, no good way of judging how the imperfections will affect the absorption coefficient. For this reason, we would regard the data obtained from the very thin samples (i.e., absorption coefficients $>10^4$) as still subject to some small uncertainty. From the lack of an annealing effect and from the cross checking of data points obtained from samples differing in thickness by about a factor of 5, it would appear that this uncertainty would be less than 20 percent in the value of the absorption coefficient.

RESULTS

The absorption spectra for germanium at room temperature and liquid nitrogen temperature are shown in Figs. 2 and 3, those for silicon in Fig. 4. Figure 2 is a plot of a portion of the germanium data on an expanded horizontal scale.

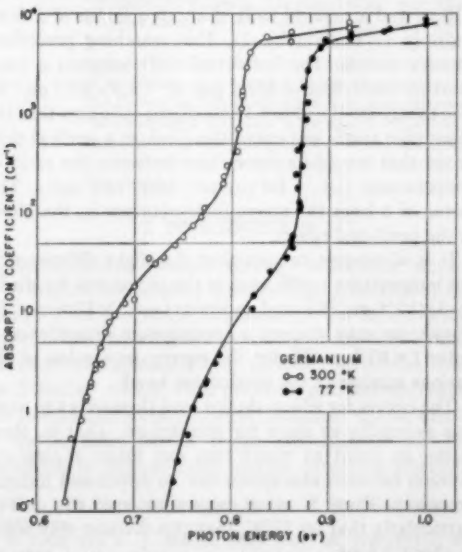


FIG. 2. Absorption spectra of single-crystal germanium at 77°K and 300°K.

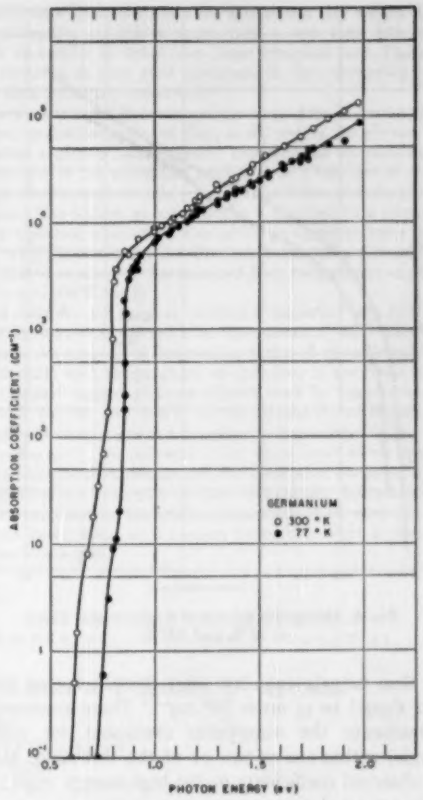


FIG. 3. Absorption spectra of single-crystal germanium at 77°K and 300°K.

It is of interest to see if the observed data can be interpreted in terms of the theoretical model suggested by Hall, *et al.*[7] Let us first consider the case for germanium. Starting at the high-energy end and going toward lower energies, there is a gradual decrease in the absorption coefficient from order 10^5 to 10^4 cm^{-1} down to an energy of 0.81 ev at room temperature and 0.88 ev at liquid nitrogen. At these points there is a very abrupt and rapid decrease in absorption coefficient from about 5×10^3 to about 10^2 cm^{-1}. (See Fig. 3.) One can interpret the absorption on the high-energy side of these "knees" as resulting largely from direct transitions and the knees themselves as indicating the approximate low-energy thresholds for these direct transitions. This view is supported by the magnitude of the absorption coefficient in the high-energy region. For example, assume that the oscillator strength for the transition is of order 3 and that the valence and conduction bands are of order 10 ev wide with a constant density of states. One

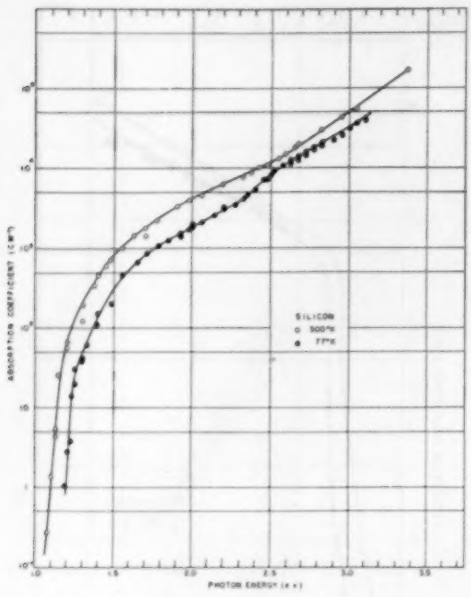


FIG. 4. Absorption spectra of single-crystal silicon at 77°K and 300°K.

can then naively calculate that the absorption coefficient should be of order 10^6 cm^{-1}.[9] These assumptions overestimate the absorption coefficient for photon energies within one or two ev of the threshold. Hence the observed coefficients in the high-energy region are in the expected range.

If we accept the premise that we are seeing direct transitions in the high-energy region, then the temperature effects on absorption coefficients in this region should, in first approximation, merely correspond to the change in band gap. That is, the curves in this range should be superposable by a horizontal shift. This is only roughly the case. There is a suggestion that the curves at the two temperatures are not strictly parallel. The horizontal shift required to bring the curves into superposition (in the direct transition region) ranges from 0.07 to 0.12 ev. The shift in the (threshold) region just below the knees is 0.07 ev. These data would indicate a temperature coefficient of band gap of from -3 to -5×10^{-4} ev/°K, the low value corresponding to the shift at the threshold.

Herman's most precise calculation of the band structure of germanium[10] indicated that the lowest states of the conduction band near $k=0$ are of S character, whereas the highest states for the valence band are of P character. If this is true, the theory for the direct transitions indicates that the energy dependence of the absorption coefficient should be as $(\nu-\nu_{\text{threshold}})^{\frac{1}{2}}/\nu$. This expression results from an application of a constant-effective-mass approximation. The observed absorption coefficient indicates a stronger energy dependence. This could result from an inapplicability of the effective mass approximation over the range of our data. In view of our knowledge of the bands, particularly the complex character of the conduction band, this seems a reasonable hypothesis. It is also possible that the lowest states of the conduction band near $k=0$ are of P character, in which case the absorption coefficient would go as $(\nu-\nu_{\text{threshold}})^{\frac{1}{2}}/\nu$. This would be a closer representation of the data but still far from a satisfactory one.

Referring to the germanium data again, it is observed that on the low-energy side of the knees, shoulders are developed in the curves at absorption coefficients of order 10^2 cm^{-1}. These are most clearly shown in Fig. 2. As the energy is further decreased the absorption coefficient again becomes a very rapidly decreasing function of energy. The absorption in the shoulders can be attributed to indirect transitions. The thresholds for these processes are presumably indicated by the regions of very rapid change of absorption coefficient with energy. As such, we would place these thresholds at ~0.62 ev and at ~0.72 ev for 300°K and 77°K, respectively.

The absorption due to indirect transitions should be temperature dependent in two ways. First, there should be a change in band gap and second a change due to an altered phonon interaction. The former process would result in a horizontal shift, the latter in a vertical shift of the absorption curves. We have found that the best superposition of the absorption curves in the range from 10^2 to 10^{-1} cm^{-1} is obtained by a horizontal shift of 0.10 ev and a vertical shift of about a factor of 2 (i.e., multiply 77°K data by 2). This matching procedure appears unique. The horizontal shift suggests a temperature coefficient of band gap of -4.5×10^{-4} ev/°K. The theory for the indirect transitions indicates that the absorption coefficient curve should show a vertical shift factor that would lie somewhere between the ratio of temperatures (i.e., 4 for present data) and unity. The factor of 2 from the above interpretation is, therefore, in the predicted range.

It is of interest to note that the slight difference in the temperature coefficients of the thresholds for direct (-3×10^{-4} ev/°K) and indirect (-4.5×10^{-4} ev/°K) transitions may suggest a temperature dependence of order 1×10^{-4} ev/°K for the energy separation of the various minima of the conduction band.

The curves for silicon do not lend themselves to analysis as readily as those for germanium. That is, there seems no point at which one can make a clear-cut division between absorption due to direct and indirect processes. There is some suggestion from the curves, particularly that for 77°K, that this division may occur at about 2.5 ev.

It is noteworthy that in the low-absorption coeffi-

[9] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, London, 1940), p. 103.
[10] F. Herman, Physica **20**, 801 (1954).

cient region it is possible to get an excellent superposition of the curves for the two temperatures by a single horizontal plus vertical shift. The matching process, which again appears unique, consists of a horizontal shift of 0.10 ev and a vertical shift factor of 1.7 (i.e., multiplying 77°K data by 1.7). This gave a superposition of curves from $\sim 5\times 10^3$ to ~ 0.5 cm^{-1} (lower limit of data). The horizontal shift gives a temperature coefficient of band gap of -4.5×10^{-4} ev/°K. The vertical shift is again of the order of magnitude predicted by theory.

ACKNOWLEDGMENTS

We are greatly indebted to L. Keifer and J. H. McTaggart for their painstaking work in the preparation of the samples. We should also like to thank A. G. Tweet for the Hall measurements on the thin samples. We have profited from discussion with E. O. Kane of the theoretical aspects of this work. We are indebted to H. Brooks for comments on the manuscript prior to publication.

Note added in proof.—Since the preparation of this manuscript, MacFarlane and Roberts [Phys. Rev. **97**, 1714 (1955); **98**, 1865 (1955)] have presented data for germanium and silicon in the low-absorption coefficient range (i.e., α less than 100 cm^{-1}). Their data differ in detail from those presented here. The latter are, however, in very good agreement in the overlapping range with data previously presented.[4,6]

MacFarlane and Roberts analyze their data in terms of the indirect transition model of Hall, *et al.*,[7] taking the phonon contribution explicitly into account. This they do by plotting the square root of the absorption coefficient as a function of photon energy and decomposing the resultant curve into two straight line sections. The section at low energies is ascribed to a process in which a phonon is absorbed, the section at higher energies to one in which a phonon is emitted. From their data they estimate that the phonon required is characterized by a temperature of 260°K for Ge and 600°K for Si.

Our data do not show as obvious a resolution into two such linear sections when plotted after the manner of MacFarlane and Roberts. An analysis of our data in terms of vertical and horizontal shifts with temperature, as described in the body of our paper, would suggest phonons characterized by temperatures of ~300°K for Ge and 500°K for Si. However, as Brooks has pointed out to us, there is a temperature dependence in the term $(E_{c_0}-E_{c_{(111)}})^{-2}$ (see reference 7 for definitions) which appears in the theoretical expression for the absorption coefficient. This dependence is comparable to the temperature dependence explicitly expressed by the phonon population. Consequently, any analysis based solely on a phonon population effect is probably not too meaningful.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Surface Barrier Analysis for Metals by Means of Schottky Deviations*

D. W. JUENKER
University of Notre Dame, Notre Dame, Indiana

(Received April 15, 1955)

The transmission coefficient for the mirror-image barrier at a metal surface is applied to the case of photoelectric emission. The distribution over which the coefficient is averaged is that given by Fowler for energies normal to the emitter surface, and is altered by the Schottky barrier lowering. The average differs from unity by an amount the major part of which is a periodic function of the accelerating field, giving a periodic deviation from the photoelectric Schottky effect which differs from the thermionic deviation only in amplitude. A refinement in the computation of the unaveraged periodic transmission coefficient brings the theory for thermionic deviations into agreement with that of Miller and Good. The improved form of the thermionic deviation is applied to experimental data to evaluate the complex reflection coefficient μ characterizing the potential form in the immediate vicinity of the surface. The values of $|\mu|$ so obtained for the highly refractory metals are of the order of 0.4, as compared with 0.2 predicted on the basis of the box model.

I. INTRODUCTION

ELECTRONS emitted through a metallic surface into an accelerating field undergo reflection at the surface itself and at the maximum of the mirror-image barrier. The interference resulting from this double reflection manifests itself in periodic deviations from the Schottky effect.[1] In a previous publication,[2] there was formulated a transmission coefficient characteristic of the metallic barrier shape and suitable for the discussion of any emission process involving electrons of low average escape energies. A particular application was made to the case of thermionic emission. In the present work, the parallel case of photoelectric emission will be considered. In the early stages of this work, it was found that the periodic part of the transmission coefficient could be expressed in better form.[3] The only effect of this refinement is to increase the theoretical amplitude of the periodic Schottky deviation; the basic conclusions found previously[2] remain unchanged. The revised transmission coefficient will be described here, and its effects on the interpretation of thermionic data will be discussed before applying it to the formulation of the photoelectric theory.

* This research was sponsored by the U. S. Atomic Energy Commission.

[1] E. Guth and C. J. Mullin, Phys. Rev. **59**, 575 (1941); **59**, 867 (1941).

[2] Juenker, Colladay, and Coomes, Phys. Rev. **90**, 772 (1953).

[3] The author is indebted to Dr. Conyers Herring for pointing out the suitable process in a private communication.

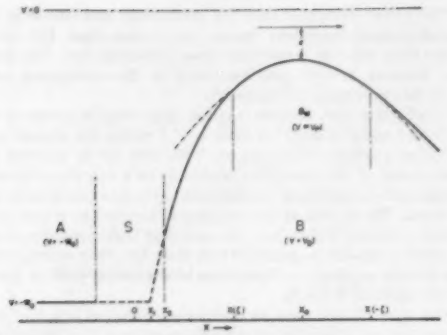


FIG. 1. Descriptive diagram of electronic potential energy *vs* distance from the metal surface. An electron escaping with excess energy ϵ passes from a nonreflecting region A, through a reflecting region S at the surface, into the mirror-image region B. Within B is a region B_M, including the barrier maximum at x_0, which is capable of reflecting electrons having low energy ϵ. In region B_M, between $x(\zeta)$ and $x(-\zeta)$, the mirror-image potential may be approximated as a parabola V_P, shown as a dashed curve. For the box model, described by the dashed curve in region S, V_B and V_A are joined in a field discontinuity at x_1.

II. REVISED TRANSMISSION COEFFICIENT

For a system of two plane parallel reflectors, the electron transmission coefficient for normal incidence may be written in terms of complex wave-reflection coefficients characteristic of the reflectors.[2] Letting these be μ and λ, and assuming $|\mu|^2 \ll 1$, the transmission coefficient has the form

$$D = 1 - |(\lambda+\mu)/(1+\lambda\mu^*)|^2 \simeq D_0 + D_1 + D_2, \quad (1a)$$

where

$$D_0 = 1 - |\lambda|^2, \quad (1b)$$

$$D_1 = -|\mu|^2[(1-|\lambda|^2)^2 - 2|\lambda|^2(1-|\lambda|^2)\cos 2\sigma], \quad (1c)$$

$$D_2 = 2|\mu||\lambda|(1-|\lambda|^2)\cos\sigma, \quad (1d)$$

$$\sigma = \arg\lambda - \arg\mu + \pi. \quad (1e)$$

Such a doubly reflecting system is presented by the metallic surface barrier, shown in Fig. 1. In the region B, beyond some distance x_2 outside the surface, an electron moves in a potential composed of the effects of its mirror image and the applied field, as given by

$$V_B = -(2x)^{-1} - x(2x_0^2)^{-1}, \quad (2)$$

where distances are in units of the first Bohr radius ($a_0 = \hbar^2/me^2 = 0.529$ A) and V_B is in units of the hydrogen ionization potential ($W_H = me^4/2\hbar^2 = 13.58$ ev). The field-dependence of V_B is embodied in the position x_0 of the barrier maximum, given by $x_0 = 3.587 \times 10^4 E^{-\frac{1}{2}}$ for the field E in v-cm^{-1}. In the vicinity of x_0, an electron having a low escape energy ϵ has some probability of being reflected, in the case of positive ϵ, or transmitted, for negative ϵ. This region is designated B_M in Fig. 1, and is bounded at $x(\zeta)$ and $x(-\zeta)$. Between x_2 and $x(\zeta)$ lies a comparatively nonreflecting region, and in region A, in the metal's interior, one can assume a plane wave function to describe the electron. Separating A and B lies the reflecting region S brought about by the surface field.

The complex wave-reflection coefficients for the regions S and B_M have been evaluated[2] and represented by the symbols μ and λ, respectively. For electrons incident on the barrier from region A one can arbitrarily assign to μ and λ the following forms, which will prove convenient in subsequent manipulations:

$$|\mu| = \cos\Delta, \quad (3a)$$

$$\arg\mu = \pi + \phi - \Delta, \quad (3b)$$

$$|\lambda| = |\lambda_0|, \quad (4a)$$

$$\arg\lambda = \arg\lambda_0 + \left(\pi + \phi' - 2\Delta - 2\int_{x_0}^{x_2} \kappa_B dx\right), \quad (4b)$$

where κ_B is the wave number of an electron moving in the potential given in (2). The quantities ϕ, ϕ', and Δ depend on the barrier shape in region S. In (4), the quantity λ_0 is the reflection coefficient to be attributed to the region B_M considered as an isolated reflector. Thus $\arg\lambda_0$ represents the phase change suffered by an electron wave reflected in B_M. The integral term in (4b) corresponds to the phase accumulation of such a wave proceeding from x_2 to its reflection in B_M and returning, and the quantity $(\pi+\phi'-2\Delta)$ may be interpreted as the phase change incurred in two opposite transits of region S. Then the quantity σ in (1e) describes the interference to the left of region S between a wave reflected in S and one reflected in B_M.

The evaluation of λ_0 proceeds as follows: from (2),

$$\begin{aligned}\kappa_B &= [\epsilon + V_B(x_0) - V_B(x)]^{\frac{1}{2}} \\ &= [\epsilon + (x_0-x)^2/(2x_0^2 x)]^{\frac{1}{2}} \\ &= [4\beta\epsilon + z^2(1-\beta^{\frac{1}{2}}z/x_0)^{-1}]^{\frac{1}{2}}/[4\beta]^{\frac{1}{2}}, \end{aligned} \quad (5a)$$

where

$$z = (x_0 - x)/\beta^{\frac{1}{2}} \quad (5b)$$

and

$$\beta = (x_0^3/2)^{\frac{1}{2}}. \quad (5c)$$

If one assumes the existence of a position $x(\zeta)$ where $z = \zeta$, such that

$$|4\beta\epsilon| \ll \zeta^2 \ll x_0^2/\beta,$$

then one can consider the potential in region B_M to be parabolic, with the corresponding wave function[2] (for a pure outgoing wave as $x \to \infty$)

$$\begin{aligned}\psi_P \propto z^{-\frac{1}{2}}&[(2\pi)^{\frac{1}{2}}/\Gamma(\tfrac{1}{2} - i\beta\epsilon)] \\ &\times \exp[-\pi\beta\epsilon/4 - i(z^2/4 + \beta\epsilon \ln z - \pi/8)] \\ &+ z^{-\frac{1}{2}} \exp[-3\pi\beta\epsilon/4 + i(z^2/4 + \beta\epsilon \ln z - 3\pi/8)],\end{aligned}$$

while a WKB wave function satisfies (5a) for $x_2 < x < x(\zeta)$:

$$\psi_B = b_1\kappa_B^{-\frac{1}{2}} \exp\left(i\int_{x_0}^{x} \kappa_B dx\right) + b_2\kappa_B^{-\frac{1}{2}} \exp\left(-i\int_{x_0}^{x} \kappa_B dx\right).$$

The joining of ψ_B to ψ_P at $x(\zeta)$ gives[2]

$$\lambda_0=b_2/b_1$$
$$=(1+e^{2\pi\beta\epsilon})^{-\frac{1}{2}}\exp i\Big[(C+\ln\zeta^2)\beta\epsilon+\zeta^2/2+2\int_{x_0}^{x(\zeta)}\kappa_B dx-\pi/2\Big], \quad (6)$$

where $C=(\gamma_E+2\ln 2)=1.96$, γ_E being Euler's constant. Using (3b), (4b), and (6) in (1e), one has

$$\sigma=(C+\ln\zeta^2)\beta\epsilon+\zeta^2/2+\pi/2-\Delta+\phi'-\phi+2\int_{x_2}^{x(\zeta)}\kappa_B dx. \quad (7a)$$

Now, throughout the integration interval of the last term of (7a), the condition $z^2\geqslant\zeta^2\gg|4\beta\epsilon|$ holds, so that κ_B may be expanded about $\epsilon=0$. Neglecting all but the first two terms, it becomes

$$\kappa_B=\kappa_0+\epsilon/(2\kappa_0),$$

where $\kappa_0=\kappa_0(x)=\kappa_B(\epsilon=0)$. Then, for $x_2\ll x_0$,

$$2\int_{x_2}^{x(\zeta)}\kappa_B dx\simeq y-(8x_2)^{\frac{1}{2}}-\zeta^2/2-(4+\ln\zeta^2-\ln y-\ln 12)\beta\epsilon, \quad (7b)$$

where

$$y=4(2x_0)^{\frac{1}{2}}/3=357.1E^{-\frac{1}{2}} \quad (8)$$

for field E in v-cm^{-1}. Replacing (7b) in (7a) and using $\alpha=|\beta\epsilon|$, one has

$$\sigma=\sigma_0\pm g\alpha, \quad (9a)$$

which is the form arrived at in previous work.[2] Now, however,

$$\sigma_0=(y+\pi/2)-\delta, \quad (9b)$$
$$\delta=\Delta+(8x_2)^{\frac{1}{2}}+(\phi-\phi'), \quad (9c)$$
$$g=C-4+\ln 12+\ln y. \quad (9d)$$

The phase angle δ, which is taken as characteristic of region S, has been changed only in formulation from the value given in reference 2. The explicit dependence on the position x_2 has been included to point out that the surface reflecting region under investigation must take in all of the unknown territory between the limits of applicability of the constant interior potential and mirror-image law. However, the major contribution to the reflection cannot be made very far outside the surface, since it is found experimentally that δ is not a function of applied field. The origin of the components of δ is given in (3) and (4).

The value of g given in (9d) differs substantially from that given in Eq. (12c) of reference 2. The latter was found by numerical solution of the phase integral

$$\int_{x_1}^{x_0}\kappa_B dx=\int_{x_1}^{x_0}\kappa_0 dx+f(\alpha),$$

as a result of which g was considered a constant of the order of 5.3 over the effective range of α. The present value of g, in (9d), takes on values between 4.0 and 2.5 for fields ranging from 10 to 2000 kv-cm^{-1}. It is preferable to the earlier form since its derivation is consistent with the other approximational methods employed in the evaluation of λ.

For the sake of comparison, one can evaluate $|\mu|$ and δ for the special case of the box model. For this example, the potential in region S takes on the particular shape shown by the dashed curve in Fig. 1. The mirror-image form is assumed valid beyond the point x_1 where it crosses the average interior potential $V_A=-W_a$. For this case, the components of μ, in (3), can be evaluated by joining at x_1 a plane wave in region A to a WKB wave in region B. This procedure gives

$$\phi_c=\phi_c'=2x_1\kappa_1\simeq W_a^{-\frac{1}{2}}, \quad (10a)$$
$$\Delta_c=\cot^{-1}|\kappa_1'/4\kappa_1^2|\simeq\cot^{-1}(W_a^{\frac{1}{2}}/4), \quad (10b)$$

where $\kappa_1=\kappa_B(x_1)\simeq W_a^{\frac{1}{2}}$ and $\kappa_1'=(d\kappa_B/dx)_{x_1}\simeq -W_a^{\frac{1}{2}}$. Thus, using (3a) and (9c), and replacing x_2 by x_1,

$$|\mu|_c=\cos\Delta_c\simeq W_a^{\frac{1}{2}}(16+W_a)^{-\frac{1}{2}}, \quad (11a)$$
$$\delta_c=\Delta_c+(8x_1)^{\frac{1}{2}}=\cot^{-1}(W_a^{\frac{1}{2}}/4)+(W_a^{\frac{1}{2}}/4)^{-1}. \quad (11b)$$

III. THERMIONIC SCHOTTKY DEVIATION

The foregoing development changes the thermionic results of reference 2 only in the average of the transmission coefficient D_2. D_2, as stated in (1d), can be constructed from the parts given in (3), (4), (6), and (9), and is given in expanded form in Eq. (18), to follow. It must be averaged over the Maxwellian distribution

$$N_t(\alpha)=K_t e^{-2\pi B\alpha}, \quad (12)$$

where

$$\alpha=|\beta\epsilon|, \quad (13a)$$
$$B=(2\pi\beta kT)^{-1}, \quad (13b)$$

and where K_t is a constant of proportionality independent of α. The averaging integration is straightforward, and results[2] in a convergent summation of terms periodic in σ_0. The summation is readily evaluated for given values of $y=357.1E^{-\frac{1}{2}}$, providing numerical values for $f_1(y)$ and $f_2(y)$ in the relation

$$\langle D_2\rangle_{Av}=|\mu|Bf_1(y)\cos[\sigma_0+f_2(y)].$$

Such a procedure has been carried out for values of y in the range $3\pi<y<12\pi$ and it is found that $f_2(y)$ varies from 0.6 by less than 0.1 radian over the whole range of y. The argument of the cosine is therefore essentially unchanged from its former value.[2] $f_1(y)$ can be represented satisfactorily by the form

$$f_1(y)\simeq 5.1y^{-0.3},$$

so that the final average of D_2 is

$$\langle D_2\rangle_{Av}\simeq 5.1y^{-0.3}|\mu|B\cos(\sigma_0+0.6). \quad (14)$$

The sizable contribution of electron tunneling near the barrier maximum is included in this expression.

Substitution of numerical values in B, as given in (13b), and multiplication by $\log_{10}e$ transposes (14) to the periodic deviation F_2 from the thermionic Schottky effect:

$$F_2 \simeq \log(I/I_0) - mE^{\frac{1}{2}} \simeq \log(1+\langle D_2\rangle_m)$$
$$\simeq 4.9\times 10^5 |\mu| T^{-1} y^{-3.2} \cos(y+2.1-\delta), \quad (15a)$$

where I_0 is the zero-field current and m the Schottky slope, and where T is the Kelvin temperature. δ is described generally in (9c) and for the special case of the box model in (11b). The small nonperiodic deviation F_1 has been neglected in (15a). By way of comparison, the result found by Miller and Good[4] contains a cosine argument nearly identical with that of (15a), and an amplitude which can be stated approximately as

$$|F_2|_{MG} = 4.6\times 10^5 |\mu| T^{-1} y^{-3.2}. \quad (15b)$$

Since the periodicity in experimental deviation data has been found to agree accurately with that of (15a), such data have been analyzed to obtain values for δ and $|\mu|$. Equation (15a) differs from the earlier value of F_2 given in reference 2 only in amplitude. Therefore the present revised form yields the same result with respect to the phase δ: experimental values for δ for the highly refractory metals are smaller by a quarter period than the values calculated from (11b), indicating the existence of a potential shape in the region S of Fig. 1 different from the field discontinuity presented by the box model. Experimental values for $|\mu|$, on the other hand, are increased in magnitude by the revision of F_2. Hence, the values of $|\mu|_s$ cited in Table II of reference 2 must be replaced by values found from (15a) above, namely,

$$|\mu|_s = 2.06\times 10^{-6} A_m T y_m, \quad (16)$$

where A_m and y_m are the amplitude and field position of the mth extremum of the experimental F_2 deviation. The values obtained for tungsten, tantalum, and molybdenum are given in Table I.

TABLE I. Amplitude analysis for thermionic data on tungsten, tantalum, and molybdenum. Here, y_m indicates the extrema analyzed, italics designating maxima. $|\mu|_s$ is computed from Eq. (16); the largest and smallest values found for $|\mu|_s$ are given. The corresponding theoretical value $|\mu|_s$ is calculated for the box model by converting the value of W_0 shown in the table to Rydberg units and substituting in Eq. (11a).

	Tungsten Seifert, Phipps[a]	Tantalum Munick, LaBerge, Coomes[b]	Molybdenum Brock, Houde, Coomes[c]
y_m	21.9, *25.3*	*24.9*, 27.9, *31.1*	*25.6*, 28.7
$\|\mu\|_s$	0.39–0.46	0.35–0.45	0.30–0.42
W_0	10.3 ev	9.3 ev	10.2 ev
$\|\mu\|_s$	0.22	0.21	0.22

[a] R. L. E. Seifert and T. E. Phipps, Phys. Rev. **56**, 652 (1939).
[b] Munick, LaBerge, and Coomes, Phys. Rev. **80**, 887 (1950).
[c] Brock, Houde, and Coomes, Phys. Rev. **89**, 851 (1953).

[4] S. C. Miller and R. H. Good, Phys. Rev. **92**, 1367 (1953).

IV. PHOTOELECTRIC SCHOTTKY DEVIATION

The transmission coefficient D will be applied to the case of surface photoelectric emission in the following development. The assumptions of Fowler[5] will be used: (a) that the number of electrons in a metal "available" for photoelectric emission is equivalent to the number whose energy component normal to the surface, augmented by the illumination energy $h\nu$, is sufficient to surmount the surface barrier; and (b) that the probability that one of these "available" electrons will be excited by illumination near the threshold is independent of energy.

For photoelectric emission, therefore, the equivalent of the Maxwellian distribution in the thermionic case is a Fermi distribution which has been integrated over the two directions parallel to the surface. This can be written

$$N_p(\epsilon) = K_p \ln[1+\exp(\chi' - \epsilon/kT)], \quad (17)$$

where

$$\chi' = h(\nu-\nu_0')/(kT)$$
$$= [h(\nu-\nu_0) - V_B(x_0)]/(kT)$$
$$= [h(\nu-\nu_0) + x_0^{-1}]/(kT),$$

ϵ is the electron energy relative to the barrier maximum as shown in Fig. 1, K_p is a constant of proportionality approximately independent of ϵ, and $h\nu_0$ is the zero-field threshold.

The transmission coefficient given in (1) can be expressed as a function of field and energy by inserting the values of λ and μ from (3), (4), (6), and (9). It can then be expanded as follows, with $\alpha = |\beta\epsilon|$:

$$D_0 = \tfrac{1}{2}(1\pm 1) \pm \sum_{n=1}^{\infty} (-1)^n e^{-n2\pi\alpha}, \quad (18a)$$

$$D_1(m) = -|\mu|^2 \left[\tfrac{1}{2}(1\pm 1) - \sum_{n=1}^{\infty} (-1)^{n+1}(n\pm 1) e^{-n2\pi\alpha} \right], \quad (18b)$$

$$D_1(p) = 2|\mu|^2 \sum_{n=1}^{\infty} (-1)^{n+1} n e^{-n2\pi\alpha} \cos 2(\sigma_0 \pm g\alpha), \quad (18c)$$

$$D_2 = 2|\mu| \sum_{n=0}^{\infty} (-1)^n \frac{\Gamma(n+\frac{3}{2})}{\Gamma(\frac{3}{2})\Gamma(n+1)} \times e^{-a_n\alpha} \cos(\sigma_0 \pm g\alpha), \quad (18d)$$

where $D_1 = D_1(m) + D_1(p)$, $a_n = 2\pi[n - (\pm\frac{1}{4} - \frac{3}{4})]$, and where the plus-or-minus signs indicate the sign of ϵ. σ_0 and g have been given in (9). For the extreme field $E = 500$ kv-cm^{-1}, $2\pi\alpha$ exceeds unity for $|\epsilon|$ greater than about 0.008 ev. Therefore, for purposes of averaging (18) over the distribution (17), $|\epsilon|$ may be considered small compared with kT, even at room temperature. It is also simpler to work with $h(\nu-\nu_0)$ of such a magnitude that χ' is large compared with $\exp(-\chi')$. At room temperature this is normally

[5] R. H. Fowler, Phys. Rev. **38**, 45 (1931).

satisfied with frequencies such that $h(\nu-\nu_0)>0.05$ ev. For this case, the logarithmic term in (17) may be expanded, retaining

$$N_p(\alpha)\simeq -K_p(\pm 2\pi B\alpha-\chi'), \tag{19}$$

where the plus and minus again refer to the sign of ϵ.

The average transmission coefficient is defined by

$$\langle D\rangle_{h\nu}=\frac{\int_0^\infty D_+(\alpha)N_+(\alpha)d\alpha+\int_0^\infty D_-(\alpha)N_-(\alpha)d\alpha}{\int_0^\infty N_+(\alpha)d\alpha}, \tag{20}$$

where the subscripts indicate the quantities in (18) and (19) evaluated for the corresponding energy ranges, the second term in the numerator being the tunnelling contribution. The denominator of (20) must be found from a two-part integration,[5] using the original form of the distribution given in (17). To a sufficient approximation it is

$$\int_0^\infty N_+(\alpha)d\alpha=\beta\int_0^{\chi' kT}N_p(\epsilon)d\epsilon+\beta\int_{\chi' kT}^\infty N_p(\epsilon)d\epsilon \simeq K_pP(4\pi B)^{-1}, \tag{21}$$

where

$$P=\tfrac{1}{3}\pi^2+\chi'^2, \tag{22}$$

and B is given in (13b). The averages of (18) can be found by applying (19), (20), and (21). Neglecting terms smaller than B^2/P and $|\mu|^2$, the results are

$$\langle D_0\rangle_{h\nu}\simeq 1+(\pi B)^2(3P)^{-1}, \tag{23a}$$

$$\langle D_1(m)\rangle_{h\nu}\simeq -|\mu|^2(1-2\chi' BP^{-1}), \tag{23b}$$

$$\langle D_1(p)\rangle_{h\nu}\simeq C_1|\mu|^2\chi' BP^{-1}\cos(2\sigma_0-\theta_1), \tag{23c}$$

$$\langle D_2\rangle_{h\nu}\simeq C_2|\mu|\chi' BP^{-1}\cos(\sigma_0+\theta_2). \tag{23d}$$

The quantities C_1, C_2, θ_1, and θ_2 have been evaluated for values of $y=357.1E^{-1}$ in the practical field range, in the manner which has been described for the thermionic case. They have been found to fit the following relationships:

$$C_1\simeq 6.16y^{-0.64};\quad \theta_1\simeq\tan^{-1}[14y^{-2.9}(\chi' kT)^{-1}];$$
$$C_2\simeq 10.12y^{-0.32};\quad \theta_2\simeq 0.6.$$

The photoelectric emission in an accelerating field should follow the relationship:

$$I=I_0[P\langle D\rangle_{h\nu}]/[P_0\langle D(0)\rangle_{h\nu}], \tag{24}$$

where $\langle D\rangle_{h\nu}$ is the sum of the averaged coefficients in (23), and where I_0, P_0, and $\langle D(0)\rangle_{h\nu}$ are the zero-field quantities. From (22) and (23), the latter two are

$$P_0=\tfrac{1}{3}\pi^2+[h(\nu-\nu_0)/(kT)]^2,\quad \langle D(0)\rangle_{h\nu}=1-|\mu|^2,$$

Discarding terms in $|\mu|^2$, one may write (24) in terms of the first-order photoelectric Schottky current I_1:

$$\begin{aligned}(I-I_1)/I_1&\simeq\langle D\rangle_{h\nu}-\langle D(0)\rangle_{h\nu}\\ &\simeq B(\pi^2B+6\chi'|\mu|^2)/(3P)\\ &\quad+C_1|\mu|^2\chi' BP^{-1}\cos(2\sigma_0-\theta_1)\\ &\quad+C_2|\mu|\chi' BP^{-1}\cos(\sigma_0+\theta_2)\\ &=\Phi_1+\Phi_2'+\Phi_2,\end{aligned} \tag{25}$$

where $I_1=I_0P/P_0$. The monatonic Φ_1, and Φ_2', which is periodic in $2\sigma_0$, have amplitudes less than 10% of that of Φ_2, even for high fields and illumination near the threshold. The quantity I_0/P_0 depends on the quantum efficiency of the surface and the intensity of illumination, and contains the factor $(kT)^2$. Hence, ideally, I_1 is given by the proportionality[1]

$$I_1\propto\tfrac{1}{3}(\pi kT)^2+[h(\nu-\nu_0)+48.3y^{-2}]^2.$$

However, due principally to the patch effect, I_1 probably will be an empirical curve-fit in practice, and Φ_1 will be inseparable from experimental data.

In terms of the field parameter y, as given in (8), and kT and $h\nu$ in electron volts, and with the values of

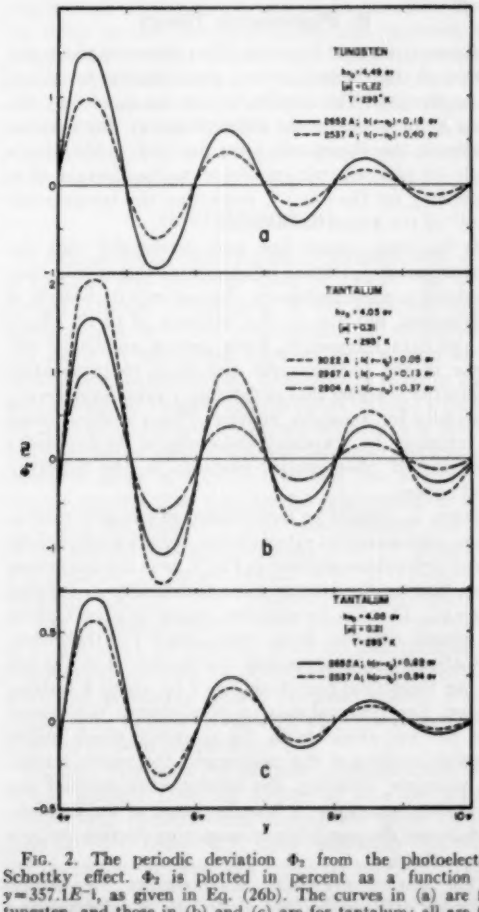


FIG. 2. The periodic deviation Φ_2 from the photoelectric Schottky effect. Φ_2 is plotted in percent as a function of $y=357.1E^{-1}$, as given in Eq. (26b). The curves in (a) are for tungsten, and those in (b) and (c) are for tantalum; all are for room temperature, with illumination wavelengths corresponding to the more useful lines of the mercury spectrum. Note the doubled ordinate scale in (c).

C_1, C_2, and θ_2 included, Φ_2' and Φ_2 are

$$\Phi_2' = \frac{126|\mu|^2 y^{-3.6}\{h(\nu-\nu_0)+48.3y^{-2}\}}{\{\frac{1}{3}(\pi kT)^2+[h(\nu-\nu_0)+48.3y^{-2}]^2\}} \times\cos[2y+\pi-\theta_1(y,\nu)-2\delta], \quad (26a)$$

$$\Phi_2 = \frac{208|\mu|y^{-3.2}\{h(\nu-\nu_0)+48.3y^{-2}\}}{\{\frac{1}{3}(\pi kT)^2+[h(\nu-\nu_0)+48.3y^{-2}]^2\}} \times\cos[y+2.1-\delta]. \quad (26b)$$

Φ_2 comprises almost the entire deviation from the photoelectric Schottky effect; it is the counterpart of the periodic deviation F_2 in the thermionic case [see Eq. (15a)]. It is represented graphically in Fig. 2 for tungsten and tantalum, using the box model values $|\mu|_c$ for $|\mu|$ [see Table I and Eq. (11a)] and the threshold values[6] 4.05 ev for tantalum and 4.49 ev for tungsten.

V. CONCLUSIONS

A. Revised Thermionic Theory

The use of the revised transmission coefficient introduces no change in the conclusions reached in previous work,[2] other than the following additional remarks:

There is a striking similarity between the value of the thermionic Schottky deviation as determined in the present work with the revised transmission coefficient, and that found by Miller and Good.[4] The latter theory was formulated by means of a refined WKB approximational method but, as in the present case, a parabolic potential form was assigned to the barrier maximum. Therefore, although the experimental values for $|\mu|$ given in Table I may be real, suspicion has not been completely removed from the parabolic approximation as being the cause of their apparently excessive magnitudes.

With regard to these values of $|\mu|_x$, it is to be noted that essentially external factors such as patch effect or surface irregularities would tend to give rise to mutually destructive Schottky deviations, resulting in low apparent values of $|\mu|$, rather than the high magnitudes actually observed. The presence of a third reflecting layer, such as one might suppose to exist in the case of a uniformly contaminated surface, would likewise lead to low measurements of $|\mu|$ under ordinary circumstances.

It has been the practice throughout the present computations to neglect terms in $|\mu|^3$ or higher order as they occur. This approximation is justified for $|\mu|\sim 0.2$, and appears to be generally satisfactory even for values of $|\mu|$ near 0.4, since the terms in question usually occur in combination with other small quantities. However, in computing the final theoretical form of the periodic deviation in both the thermionic and photoelectric cases, the approximation

$$[\langle D(0)\rangle_{Av}+\langle D_2\rangle_{Av}]/\langle D(0)\rangle_{Av} = [1-|\mu|^2+\langle D_2\rangle_{Av}]/[1-|\mu|^2] \simeq 1+\langle D_2\rangle_{Av}$$

has been used, in keeping with the prior practice. In this instance, $|\mu|^3$ is not necessarily negligible in comparison with unity, but the retention of higher order terms lowers the values of $|\mu|_x$ in Table I by only $\sim 15\%$.

The agreement in phase and period of the periodic deviation F_2 found here and in reference 2, with that of Miller and Good, is reassuring. There was some doubt whether the ordinary WKB wave function could be validly applied at the barrier corner x_1 of the box model. Were such doubt well founded, it would seem that the improvement in method employed by Miller and Good should result in a phase shift in F_2; such a shift is not found.

B. Photoelectric Theory

A comparison of (26b) with (15a) shows the phase and period of the thermionic and photoelectric deviations to be identical. The amplitudes are also essentially the same in form; due to the different energy distributions involved, the thermionic form has a T^{-1}-dependence while the photoelectric amplitude can be thought of as depending on $(h\nu-h\nu_0)^{-1}$, neglecting the temperature "tail" of the Fermi distribution.

It has been noted here and previously[2] that the evaluation of $|\mu|$ from experimental thermionic data is subject to some ambiguity, due not only to theoretical deficiencies, but also to the influence of patch effects on the data themselves. Even greater ambiguity will occur in the photoelectric case since to these same difficulties is added that of defining a value for $h(\nu-\nu_0)$, especially for a patchy emitter. This condition could be improved by a working knowledge of the first-order behavior of photoelectric emission in the Schottky field region.

Since $|\mu|$ should be independent of emission mechanism, one can expect experimental deviation amplitudes about twice those graphed in Fig. 2, as in the thermionic case, barring the aforementioned difficulty in defining $h(\nu-\nu_0)$. There is no apparent reason to expect phase or period different from those found for thermionic deviations, i.e., experimental curves shifted to the left of the theoretical curves of Fig. 2 by about a quarter period. Experimental data on photoelectric deviations[7] are not yet available in the quantity which makes possible analysis of the counterpart thermionic results. In principle, however, the photoelectric method has obvious advantages: a broader range of subject materials, and the possibility of examining discrete portions of a surface.

The author gratefully acknowledges the helpful criticism and suggestions of Dr. E. A. Coomes and Dr. R. H. Good.

[6] C. Herring and M. H. Nichols, Revs. Modern Phys. **21**, 185 (1949), Table IV.

[7] Buder, Ruddick, and Weber, Phys. Rev. **91**, 561 (1953).

Electrical Behavior of Barium Titanate Single Crystals at Low Temperatures

H. H. WIEDER
United States Naval Ordnance Laboratory, Corona, California
(Received January 17, 1955; revised manuscript received March 15, 1955)

Barium titanate single crystals have electrical characteristics in the orthorhombic and rhombohedral phases which are similar to those observed in the tetragonal region. At a phase transition the coercivity, determined from 60 cps hysteresis loops, decreases sharply from its value at temperatures above the transition. Correspondingly, the polarization reversal time and the spontaneous polarization decrease discontinuously from the values above the transition. All of these phenomena are accompanied by thermal hysteresis. A plot of the dielectric constant as a function of switching time is a straight-line relation and supports the theory of a relaxation-type mechanism for polarization reversal.

I. INTRODUCTION

THE electrical and optical properties of barium titanate single-domain crystals have been the subject of many investigations. Recently published papers have clarified the nature of the transition from the cubic to the tetragonal phase and explored the polarization reversal as a function of various parameters in the tetragonal region. This paper describes the electrical behavior in the orthorhombic and rhombohedral phases of some $BaTiO_3$ crystal plates which contained only parallel or antiparallel domains (c-domains) at room temperature.

II. EXPERIMENTAL

A total of ten crystals ranging in thickness from 0.25 to 0.05 mm with electrode areas of 15×10^{-7} m^2 were studied. Optical observation of the electrode areas indicate these to consist of c-domains and no attempt was made to distinguish between parallel and antiparallel domain orientation. Electrodes of silver or gold were applied by evaporating the metal through a masking aperture on the selected crystal surfaces. The results described below are characteristic of the crystals in the thickness range of 0.25 to 0.10 mm, but not of the thinner crystals. The spontaneous polarization at plus 25°C is of the order of 26×10^{-6} coulomb/cm^2 and the 60-cps coercivity is approximately 1.2×10^3 volts/cm. The behavior of these crystals (0.25 to 0.10 mm in thickness) appeared to be completely reproducible. The thinner crystals, however, when brought below minus 20°C appeared to develop more or less permanent domain reorientations, i.e., wedge-shaped a-domains (in what had previously been a c-domain region) which persisted for days or weeks after the crystal had been returned to room temperature.

Figure 1 illustrates the 60-cycle hysteresis loops of a crystal brought from the tetragonal into the orthorhombic and rhombohedral phases. The phase transitions at −5°C and −92°C agree reasonably well with the transition temperatures found by others.[1–3] Temperatures were recorded by means of a thermocouple in contact with the body of the crystal and the circuit employed for the display of the loops is the—by now standard—Sawyer and Tower arrangement.

Figure 2 shows the coercivity as a function of temperature calculated from the hysteresis loops of Fig. 1 and others not shown. Particularly noteworthy is the sharp decrease in coercivity in the vicinity of the phase transitions. Figure 3 shows the spontaneous polarization P_s as a function of temperature obtained from the same hysteresis loops. The agreement with the theoretical curves of Devonshire[4] is reasonably good. All crystals

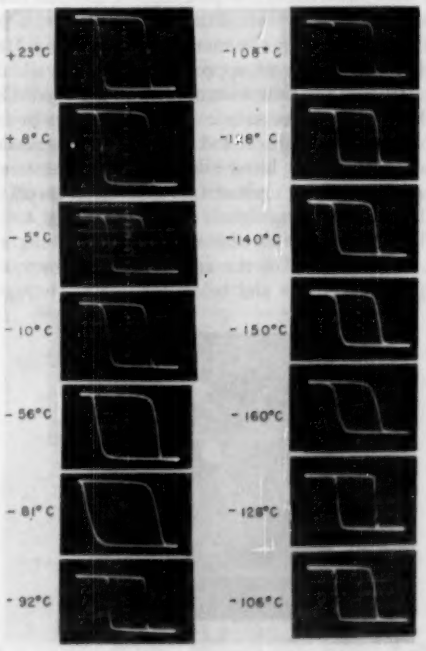


FIG. 1. Hysteresis loops as a function of temperature for the c-domain area of a $BaTiO_3$ crystal. For all loops $E=0.63$ kv/cm per horizontal division and $Q=0.11\times10^{-6}$ coulomb per vertical division.

[1] von Hippel, Breckenridge, Chesley, and Tizza, Ind. Eng. Chem. **38**, 1097 (1946).
[2] H. F. Kay and P. Vousden, Phil Mag. 7, **40**, 1019 (1949).
[3] W. J. Merz, Phys. Rev. **76**, 1221 (1949).
[4] A. F. Devonshire, Phil. Mag. 7, **40**, 1050 (1949).

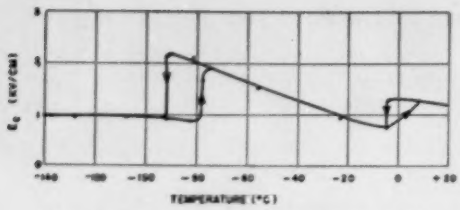


FIG. 2. Coercivity as function of temperature in the orthorhombic and rhombohedral phases.

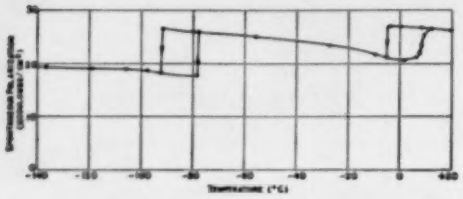


FIG. 3. Spontaneous polarization as a function of temperature.

without exception exhibited the sharp decrease in coercivity following a phase transition. There was a considerable variation, however, in the curves of P_s as a function of temperature between individual crystals.

A determination of the polarization reversal time τ_s as a function of the amplitude of the applied field E was made by using the method described by Merz.[5] Rectangular pulses of approximately 6-μsec duration and 0.04-μsec rise time were applied to the crystal from a 50-ohm impedance source. The crystal was in series with a 25-ohm resistor and the switching transients obtained across the latter allowed the measurement of the peak switching current i_{max} and the reversal time τ_s. Typical switching signals are shown in Fig. 4.

Figure 5 illustrates the reciprocal switching time $1/\tau_s$ as a function of the applied field E for various temperatures above and below the -5°C (tetragonal to orthorhombic) transition. For values of $E>5E_c$ the linear relation between $1/\tau_s$ and E was found valid for all crystals and therefore, in this region, τ_s would appear to be related to the propagation of antiparallel domains as described by Merz for the tetragonal region. For values of $E<5E_c$, considerable variations in the plot of $1/\tau_s$ *versus* E were noted and a report on this behavior will be deferred.

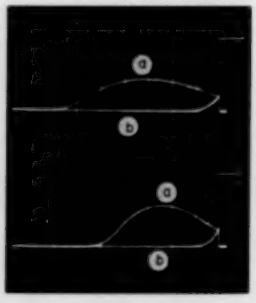


FIG. 4. Representative switching transients obtained by using a rectangular pulse of 6-μsec duration of positive or negative sign and 8.8 kv/cm amplitude: *a* indicates reversal, *b* nonreversal of polarization. Upper photograph taken at +23°C, abscissa = 0.3 μsec/div; lower photograph taken at -18°C, abscissa = 0.1 μsec/div. Ordinate for both is = 0.29 amperes/division.

[5] W. J. Merz, Phys. Rev. **95**, 690 (1954).

Figure 6 is a plot of the reciprocal switching time as a function of temperature for a constant field of $E=8.8$ kv/cm. All crystals gave analogous curves to Fig. 6 and, as was to be expected from the linear relations of $1/\tau_s$ *vs* E, the characteristic shape of the curve is independent of E for $E\gg E_c$.

Figure 7 shows the peak switching current, i_{max}, as a function of E for various temperatures in the orthorhombic and rhombohedral phases. In his investigations of the polarization reversal in the tetragonal phase, Merz[5] derives a switching resistance: $R=\beta d^2/2qAP_s$, where A and d refer to the electrode area and thickness of the crystal, P_s is the value of spontaneous polarization of a single domain, and q is a shape factor determined from the $i(t)$ curves. The slope of the line $1/\tau_s$ *versus* E is proportional to the domain mobility factor β.

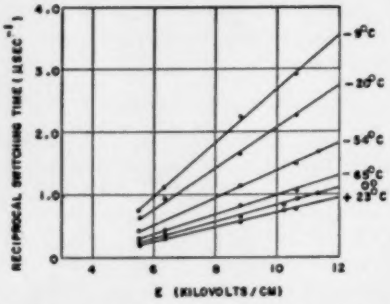


FIG. 5. Reciprocal switching time, $1/\tau_s$, as a function of field E and temperature T in the orthorhombic phase.

Since the crystal used in these experiments had an area of 10^{-2} cm^2 and thickness of 2×10^{-2} cm, compared to the crystal described by Merz ($A=10^{-4}$ cm^2, $d=5\times10^{-3}$ cm, $R=500\Omega$), the ratio of the resistances at room temperature should be of the order of 1:6. The value of $R=290$ ohms obtained from the slope of the $i(t)$ *vs* E curve of Fig. 7 is higher than expected but in view of possible large differences within crystals, not unreasonable.

III. DISCUSSION

The curve of $P_s(T)$ shown in Fig. 3 is in the orthorhombic phase in qualitative agreement with Devonshire's theory[4] and his computations. Based on crystal thermodynamics and the experimentally determined changes in the lattice structure of the $BaTiO_3$ unit cell, this theory accounts for the three phase transitions, the behavior of the initial dielectric constant as a

function of temperature, and other experimentally verifiable properties of the single-domain barium titanate crystal. The discontinuous jump in P_s at the ferroelectric transition (−5°C) from the tetragonal to the orthorhombic phase is in close agreement with the theory: P_s (orthorhombic)$=\sqrt{2}P_s$ (tetragonal). This agreement is lacking however for the lower (−92°C) transition, possibly because the crystal is no longer a c-domain. Latter observations made by means of a polarizing microscope indicated complicated twinning effects occurring in the rhombohedral phase.

It should be emphasized that a closer agreement between theory and experiment was obtained by Kay and Vousden[2] for smaller absolute values of spontaneous polarization ($P_s=16\times10^{-6}$ coulomb/cm² at room temperature). The crystals used for the experiments described in this paper have higher values ($P_s\simeq26\times10^{-6}$ coulomb/cm²) in agreement with recent measurements described by Merz.[6] It is expected, nevertheless, that complete agreement with the theory would result if a crystal were to retain a single-domain identity over the entire ferroelectric temperature range.

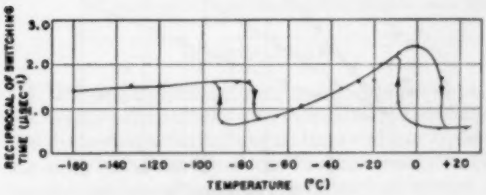


FIG. 6. $1/\tau_s$ as a function of temperature for constant field $E=8.8$ kv/cm.

A qualititative interpretation of the curve $E_c(T)$ shown in Fig. 2 can be obtained by using the graphical representation of the Devonshire equation for the tetragonal phase used by Merz[6] to indicate a theoretical relation between an idealized hysteresis loop and the double hysteresis loop observed above the normal Curie temperature. For convenience, the graph is reproduced in Fig. 8. Note the p and e are a normalized polarization and electric field respectively, and t is a temperature dependent parameter. The sign of t is positive above and negative below the Curie temperature. Points a and a' define a threshold field where the negative going slope and consequent instability causes a jump in polarization to a new, stable value. If this threshold field is tentatively identified with a coercive field, it becomes evident that for negative values of t the coercivity increases. Similar graphs can be obtained for the orthorhombic and rhombohedral phases and it is found that in agreement with the experimental observations shown in Fig. 2, the coercivity decreases for a rise in temperature as the crystal approaches a phase transition.

The data presented in Fig. 5 suggest that for the orthorhombic phase, in analogy to Merz's[6] expression,

[6] W. J. Merz, Phys. Rev. **91**, 513 (1953).

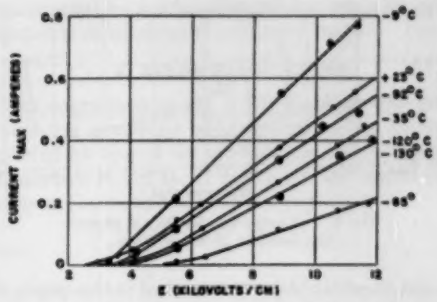


FIG. 7. Peak switching current i_{max} as a function of field and temperature.

the switching time is given by

$$\tau_s=\beta d/(E-E''),$$

where β is the reciprocal of the antiparallel domain mobility as defined by Merz and the applied field E is large enough so that only the propagation of the domains has to be taken into consideration, the nucleation time of these domains being negligible in comparison. E'' is the extrapolated value of $1/\tau_s$ for $\tau_s\to\infty$. Figure 9 illustrates the temperature dependence of β for a constant applied field. It is evident that at a ferroelectric transition the domain mobility $1/\beta$ increases discontinuously from its value above the transition. The mobility gradually decreases as the temperature is lowered and the crystal symmetry is decreased. The behavior of the dielectric constant as a function of temperature for a c-domain crystal had been reported by Merz.[3] It is reproduced for convenience in Fig. 10. A comparison with the curve of $1/\tau_s$ as a function of T of Fig. 6 shows strong similarities. If the dielectric constant is plotted as a function of $1/\tau_s$ for the tetragonal phase as reported by Merz, and for

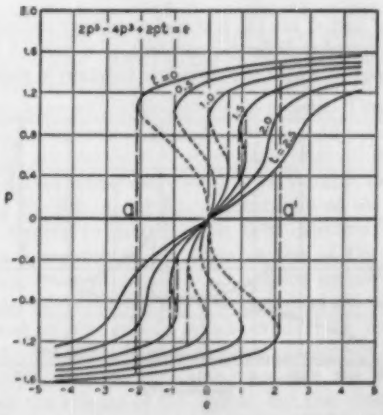


FIG. 8. The plot of a normalized equation in E, P, and T, after W. J. Merz, [Phys. Rev. **91**, 513 (1953)] for temperatures above the normal Curie point.

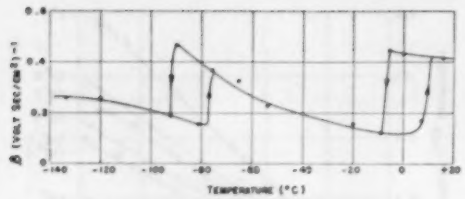


FIG. 9. The reciprocal mobility β, plotted as a function of temperature.

the orthorhombic phase as described in this paper, the straight line relationships indicated in Fig. 11 result.

Young and his co-workers[7] have measured the dielectric constant and dissipation factor of a single crystal of $BaTiO_3$ as a function of frequency. They found that the results thus obtained can be fitted by a Debye relaxation spectrum for the dielectric constant of the type

$$\epsilon^*=\frac{\alpha}{1+j\omega\tau}+\epsilon_\infty,$$

where ϵ^* is the complex dielectric constant, τ the relaxation time, ϵ_∞ is the extrapolated dielectric constant for very high frequencies, and α the polarizability of a single domain. For applied fields of the order of the coercive field, the calculated value of $\tau\simeq 5.5\ \mu$sec was found to agree well with the observed polarization reversal time of a single domain. Merz had already indicated that antiparallel domain nucleation occurs in $BaTiO_3$ rather than domain wall motion. If the assumption is now made that for a single domain there is a field dependent distribution of relaxation times and that the reversal of polarization is approximated by a rotation, one may write

$$\epsilon'=\epsilon_\infty+\frac{\alpha}{1+\omega^2\tau^2}\qquad\epsilon''=\frac{\alpha}{1+\omega^2\tau^2}\omega\tau.$$

Here, ϵ' and ϵ'' are the real and imaginary components of the complex dielectric constant, ω is related to the frequency at which the dielectric constant is measured

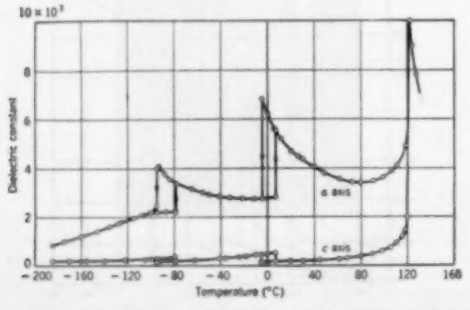


FIG. 10. Initial dielectric constants of barium titanate as a function of temperature [after W. J. Merz, Phys. Rev. **76**, 1221 (1949)].

[7] Drougard, Funk, and Young, J. Appl. Phys. **25**, 1166 (1954).

and $1/\omega\gg\tau$. Therefore

$$\frac{1}{\tau}\left(\frac{\epsilon''}{\omega}\right)=\epsilon'-\epsilon_\infty.$$

A plot of ϵ' as a function of $1/\tau_s$ as determined from curves of ϵ *versus* T and $1/\tau_s$ *versus* T should yield a straight line having the intercept $\epsilon'=\epsilon_\infty$. The value of ϵ_∞ determined from Fig. 11 is $\epsilon_\infty\simeq 160$ which compares with a value of $\epsilon_\infty=210$ to 290 calculated by Young *et al.*[7]

IV. SUMMARY AND CONCLUSIONS

A $BaTiO_3$ crystal which is a *c*-domain at room temperature will rarely maintain this quality over a wide temperature range in the lower symmetry ferroelectric phases. The measurements described are, therefore, to be considered of a preliminary nature. Certain con-

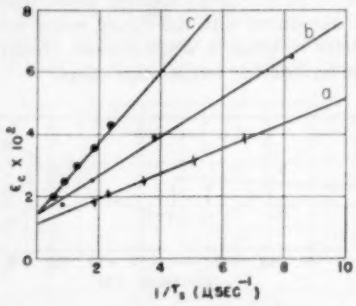


FIG. 11. Real part of complex dielectric constant E, as a function of the reciprocal switching time $1/\tau_s$. Lines *a* and *b* are for the tetragonal phase (using $1/\tau_s$ values after W. J. Merz [Phys. Rev. **95**, 691 (1954)]. Line *c* is for orthorhombic phase (from data presented in this paper). *a* taken at 4 kv/cm, *b* at 6 kv/cm, and *c* at 9.9 kv/cm.

clusions, however, may be drawn on the basis of the experimental evidence presented in this paper:

1. Within each ferroelectric phase, the electrical behavior of $BaTiO_3$ crystals shows strong similarities.

2. At a phase transition, from higher to lower crystal symmetry, the spontaneous polarization decreases roughly in accord with the Devonshire theory.

3. For high values of applied field E, the reciprocal of the polarization reversal time increases linearly with E as in the case of the tetragonal phase.

4. The mobility of antiparallel domains jumps discontinuously at a transition to a value higher than that found below the transition.

5. For high values of E, the switching current is a linear function of E over most of the orthorhombic phase. However, since considerable variations were found between crystals, more work will be required in order to define the behavior of the switching resistance in the lower symmetry phases.

6. The coercivity, determined from 60-cps hysteresis loops, drops sharply at a transition only to rise again as

temperature decreases below the transition. A mathematical model of an idealized hysteresis loop based on the Devonshire equations provides a qualitative interpretation of the relation between coercivity and temperature.

7. The relation between the dielectric constant ϵ_c, and the switching time is shown to be linear for both the tetragonal and the orthorhombic phases. Support is therefore provided for the theory describing polarization reversal as a relaxation mechanism with a field dependent distribution of relaxation times.

V. ACKNOWLEDGMENTS

The suggestions of Dr. F. S. Atchison and W. W. Davis are gratefully acknowledged. The author is indebted for four of the crystals used in the preceding experiments to Dr. D. R. Young and the International Business Machines Research Laboratory.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Pressure Dependence of de Haas-van Alphen Parameters in Bismuth

W. C. Overton, Jr., and Ted G. Berlincourt
United States Naval Research Laboratory, Washington, D. C.

(Received April 4, 1955)

The Hall coefficient, A, of a bismuth single crystal has been measured at 4.2°K, in magnetic fields up to 12 kilogauss, and under liquid helium pressures up to 120 atmospheres. The period (β/E_0) of the oscillations in reciprocal field, the phase, and the monotonic part of the magnetic field dependence of A were found to depend on the pressure with small, but nevertheless easily measurable, pressure coefficients. Only the single-period term observable in bismuth when the magnetic field is parallel to the trigonal axis was studied.

INTRODUCTION

THE low-temperature oscillatory variation of magnetic susceptibility with reciprocal magnetic field, first discovered in bismuth by de Haas and van Alphen,[1] has since been found in about 14 other metals.[2–4] The magnetoresistance and Hall coefficient[5] of several of these metals are known to have similar oscillatory characteristics at low temperatures. Recent experimental studies[6–8] on some of these metals have shown that the period of the oscillations is the same for magnetoresistance and Hall coefficient as for susceptibility. The theory for the behavior of the susceptibility advanced by Peierls,[9] Blackman,[10] and Landau[11] has been found by Shoenberg[12] and others[13] to describe fairly well the observed orientation-dependent susceptibility of several metals. The "effective" electrons, upon which the field dependent susceptibility variations depend, are presumably those whose energy surfaces overlap in **k**-space from one Brillouin zone to another. The energy surfaces are presumed to be approximately hemiellipsoidal in shape where they overlap the zone faces. Since it has been observed for some metals that the periods of oscillations are the same for susceptibility, magnetoresistance, and Hall coefficient, it is believed that the same "effective" overlapping electrons (and/or underlapping holes) participate in all three phenomena, which are described by an equation of the form,

$$y=\sum_j \{f_{0j}+a_{0j}\sin(2\pi E_{0j}/\beta_j H+\theta_j)\}. \qquad (1)$$

Here y is either χ (magnetic susceptibility), R (magnetoresistance), or A (Hall coefficient), $\beta=eh/m^*c$ (m^* = the appropriate effective mass), E_{0j} is the energy value of the highest filled overlapping energy surface measured from the energy value at the bottom of the pertinent zone, and θ_j is the phase of the oscillations. The theories provide a fixed value of $\theta_j=\pi/4$ but this is not observed experimentally in general. The periods of the oscillations in Eq. (1) are the β_j/E_{0j} and these periods are determined from experimental measurements since theory does not provide suitable values for β and E_0. For an arbitrary orientation of the crystal with respect to the magnetic field direction one observes beats due to the superposition of the various oscillatory terms in Eq. (1), but for certain special orientations oscillations of a single period are observed.

Many of the metals which exhibit the de Haas-van Alphen effect have high elastic anisotropy so that the application of hydrostatic pressure not only decreases the volume of the crystal, but also changes the c/a ratio. For example, using Bridgman's[14] values for

[1] W. J. de Haas and P. M. van Alphen, Leiden Comm. No. **212A** (1930).
[2] D. Shoenberg, Physica **19**, 791 (1953).
[3] Jules A. Marcus, Phys. Rev. **71**, 559 (1947).
[4] Verkin, Lazarev, and Rudenko, J. Exptl. Theoret. Phys. (U.S.S.R.) **20**, 995 (1950).
[5] Laird C. Brodie, Phys. Rev. **93**, 935 (1954); Reynolds, Leinhardt, and Hemstreet, Phys. Rev. **93**, 247 (1954).
[6] P. B. Alers and R. T. Webber, Phys. Rev. **91**, 1060 (1953).
[7] T. G. Berlincourt and J. K. Logan, Phys. Rev. **93**, 348 (1954).
[8] Reynolds, Hemstreet, Leinhardt, and Triantos, Phys. Rev. **96**, 1203 (1954).
[9] R. Peierls, Z. Physik **81**, 186 (1933).
[10] M. Blackman, Proc. Roy. Soc. (London) **A166**, 1 (1938).
[11] L. D. Landau, see appendix of D. Shoenberg, Proc. Roy. Soc. (London) **A170**, 341 (1939).
[12] D. Shoenberg, Trans. Roy. Soc. (London) **245**, 1 (1952).
[13] T. G. Berlincourt, Phys. Rev. **82**, 242 (1952); Phys. Rev. **94**, 1172 (1954).
[14] P. W. Bridgman, Proc. Natl. Acad. Sci. **10**, 411 (1924).

the elastic compliances of bismuth, one finds the percentage decrease in length of the c-axis to be about twice that of the a-axis so that the c/a ratio is in fact changed by a very small amount. At the same time, the percent decrease in volume is about three times that of the c/a ratio. Thus the application of hydrostatic pressure not only increases the volume in **k**-space of the Brillouin zones, but slightly changes the zone shape as well.

Jones[15] has found for bismuth that the total overlap at about 300°K is very small (corresponding to about 10^{-3} electron per atom) while Shoenberg's de Haas-van Alphen measurements give approximately 10^{-5} electron per atom at low temperature, so that in any event the overlapping energy surfaces cover only a small fraction of the area in **k**-space of the zone faces. It is not expected that the shape and orientation of the overlap surfaces will change significantly due to pressure since there is a considerably smaller relative change in zone shape compared to the change in zone volume for the "small" high pressures (up to 150 atmos) usually available in the laboratory. Thus, in the absence of information to the contrary, one would expect appreciable effects principally from the change in E_0 due to pressure. When pressure is applied, the outward motion in **k**-space of the zone face tends to decrease the small overlap energy value, E_0, while at the same time the increase in Fermi energy and the energy values of the underlying energy surfaces due to pressure tends to compensate the change in E_0. In real crystals, a change in E_0 would be expected, but it is difficult to know in advance whether to expect an increase or decrease or whether any change occurring would be large enough to manifest itself in perceptible changes in the period of the de Haas-van Alphen oscillations.

These qualitative considerations led us to look for a change in the period of oscillation due to pressure. Experimentally, it would be very difficult to study the pressure dependence of the susceptibility oscillations, so we chose to make the more convenient Hall effect and magnetoresistance measurements. Bismuth is perhaps the most suitable metal for such exploratory experiments because it exhibits very large amplitude oscillations.

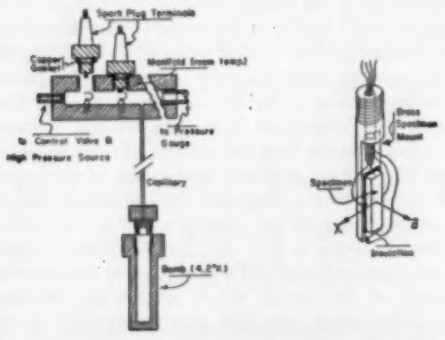


FIG. 1. Cross-section view of the high-pressure manifold and pressure bomb and a sketch (enlarged) of the specimen mount.

EXPERIMENTAL

The experimental setup shown in Fig. 1 is similar to that used by Sizoo and Kamerlingh Onnes[16] in their studies of the pressure dependence of superconductivity. The brass pressure bomb, filled with liquid helium under high pressure, contains the metal specimen mounted on its holder. The whole is situated in the bottom of a Dewar containing liquid helium at saturated vapor pressure and placed between the pole pieces of the A. D. Little Company (ADL) electromagnet. The bomb is connected by a thick-walled capillary and suspended from a high pressure manifold (at room temperature) at the top of the Dewar. The electrical leads are fed from the low-temperature—high-pressure region inside the bomb up the capillary, into the high-pressure manifold, and thence to the exterior via ordinary automobile spark plugs.

The bismuth crystal was grown from 99.99 percent pure bismuth by the Shubnikow technique in the shape of a bar of width, w, and thickness, t. It was seeded so that the trigonal axis was perpendicular (within a few degrees) to the length-width face of the bar and with a binary axis perpendicular to the length-thickness face. This crystal was mounted on the specimen mount as shown in Fig. 1 (enlarged with respect to the bomb). The specimen mount screws into the bomb's top cap.

The magnetic field direction was set parallel to the Z-axis (trigonal axis) within a few degrees. The current leads were attached as shown to the top and bottom of the specimen and provide for a current flow in the V-direction. The Hall potential leads were attached, as nearly opposite each other as possible, to the centers of the length-thickness faces. Since it is difficult to attach these Hall potential leads exactly opposite each other, a small magnetoresistance component always develops. To eliminate the magnetoresistance component, we followed the usual practice of reversing the magnetic field so that the Hall potential is then $\frac{1}{2}\{V(H)-V(-H)\}$ while the term $\frac{1}{2}\{V(H)+V(-H)\}$ permits evaluation of the interfering magnetoresistance component.

In Fig. 1, the left side of the high-pressure manifold was connected to a high-pressure valve and thence, via thick-walled beryllium copper tubing, to a pressure regulator (not shown). The regulator was a conventional helium reduction valve modified so that the beryllium copper pressure line connected to the high-pressure side of the gauge. An additional high-pressure gauge was mounted on the high-pressure manifold itself. The high pressure was always built up in a slow step-wise

[15] H. Jones, Proc. Roy. Soc. (London) **A147**, 396 (1934).

[16] G. J. Sizoo and H. Kamerlingh Onnes, Leiden Comm. No. **180b** (1926).

manner by alternately opening and closing the valve on the high-pressure helium cylinder and the valve on the high-pressure manifold. A similar slow step-wise procedure was used in reducing the bomb pressure to 5 atmos. Such careful procedures reduce the possibility of mechanical shock to the specimen.

RESULTS

Several preliminary runs were made using an $X-Y$ recorder so as to plot out the Hall voltage automatically as the magnetic field was varied. The vertical displacement of the recording paper was automatically produced by the value of the Hall potential while the recording pen was moved horizontally by an amount proportional to the magnet current. Figure 2 is typical of the several preliminary runs. To obtain the results shown in Fig. 2 three field sweeps were made in the following sequence; one at 180 psi (one of the two lower curves), one with the pressure at 1700 psi (the upper faint curve), and one with the pressure back at 180 psi. In Fig. 2, the last curve at 180 psi closely overlays the first one attesting to the reproducibility of the data. Such reproducibility at low pressures was not obtained unless cold working effects were removed by at least one previous application of the high pressure prior to the 3-curve cycle discussed. Subsequent high-pressure runs followed by low-pressure runs also gave accurately reproducible high-pressure results, showing that the cold-working was essentially complete after just one application of high pressure. The magnitude of the shift in the curves due to the pressure was always considerably greater than the magnitude of the preliminary shifts due to cold-working effects. From a visual examination of the preliminary data in Fig. 2,

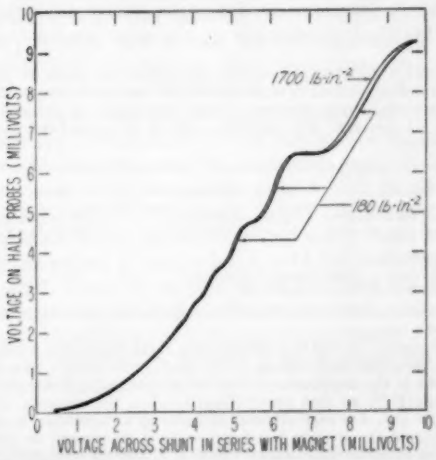


FIG. 2. Magnetic field dependence of Hall voltage in bismuth at 4.2°K and at 180 psi (12.3 atmos) and 1700 psi (116 atmos) as observed on an $X-Y$ recorder. Two curves (heavy line) at 180 psi closely overlay while the upper curve (faint line) at 1700 psi is displaced upward and to the left showing the effect of pressure on the Hall voltage.

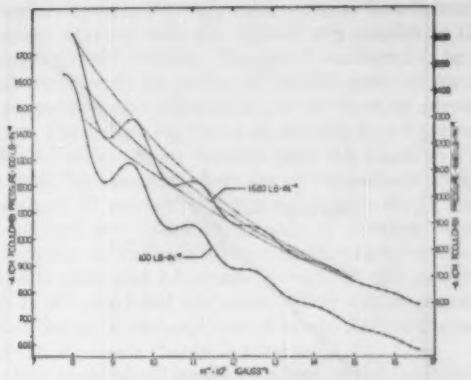


FIG. 3. The Hall coefficient in bismuth at 4.2°K *versus* reciprocal magnetic field. The field direction is parallel to the trigonal axis. The heavy curve at 1680 psi (114 atmos) is plotted with respect to the right-hand ordinate while the 100-psi (6.8 atmos) curve is plotted with respect to the left-hand ordinate. The envelopes are represented by the upper and lower dashed curves while the monotonic part of the Hall coefficient is the middle dashed curve.

a shift in phase of the oscillations is evident but a shift in the period is not easily discernible with certainty.

I. Changes in Period and Phase

It was the aim of a second series of experiments to determine the pressure effect more precisely so as to enable a positive evaluation of the changes in period and phase of the oscillations. Several precision runs were made in which the magnet current, specimen current, and Hall potential were measured potentiometrically as described above. Referring to previous field calibration data obtained with the proton resonance fluxmeter we could set the current of the ADL electromagnet so that the values of H^{-1} were convenient numbers (note the abscissa values of the experimental points in Fig. 3). The magnetic field values were accurate to better than 0.05 percent while the Hall potentials were accurate to about 0.01 percent.

The experimental value of the Hall coefficient in cm³/coulomb is $V_A t/IH\times 10^8$, where $2V_A$ in volts is $\{V(H)-V(-H)\}$, t is the crystal thickness (0.127 cm in the Z-direction in Fig. 1), I is the specimen current in amperes, and H is the magnetic field value in gauss. We found the Hall coefficient to be negative in sign (a "normal" Hall effect) for the particular orientation studied presumably signifying a predominance of electrons over holes.

In Fig. 3, typical experimental curves of the Hall coefficient are plotted *versus* H^{-1}. The lower curve representing the low-pressure results at 100 psi (6.8 atmos) is plotted with respect to the ordinate scale at the left of Fig. 3 while the upper curve for the 1680-psi (114-atmos) data is plotted with respect to the ordinate scale to the right. These two typical sets of results were purposely shifted vertically in order to separate the

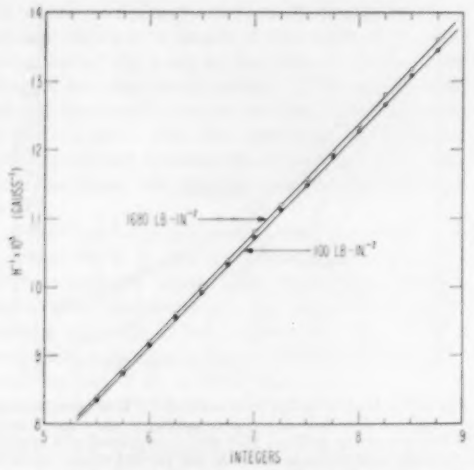


FIG. 4. Values of H^{-1} for which maxima, minima, and points of intersection with the monotonic part in the Hall coefficient occur, plotted *versus* their corresponding quarter integer values. The slopes give the period of Hall coefficient oscillations in units of H^{-1}. The open circles represent experimental data at 1680 psi and the solid circles the data at 100 psi. The heavy lines were fitted to these points by least squares. The effect of pressure is clearly shown by the separation and by difference in the slopes of the two curves.

experimental points (solid circles) which would otherwise overlap at large H^{-1} values. The envelope curves (amplitude function a_y in Eq. (1) and the upper and lower dashed curves in Fig. 3) and the monotonic part of the Hall coefficient (function f_y in Eq. (1) and the middle dashed curve in Fig. 3) were obtained by careful graphical construction. Such graphical construction can be quite accurate provided there is only one oscillatory period present. The H^{-1} values at each maximum and minimum of the oscillations picked off large-scale graphs have uncertainties of only a few parts per thousand. The H^{-1} values at the intersections of the monotonic part with the solid curve in Fig. 3 were also obtained graphically.

The usual procedure for finding the period of the oscillations is to plot the H^{-1} values at the points of the intersection of f_y with A (or independently at the minima or maxima) *versus* integers. However, if the integer scale is divided into quarter-integer values, the H^{-1} values at maxima, minima, and points of intersection can all be plotted on the same graph and the slope of the line so obtained will give a more accurate value for the period. Figure 4 is such a plot for which the solid lines have actually been calculated by least squares.[17] The slope and intercept of the line with the integer axis obtained by this calculation provide probably the best possible values for the period and phase of the oscillations. At the same time, the calculations provide the probable errors in period and phase (we assumed equal weights for the values of the maxima, etc.).

Table I shows the period and phase with their probable errors for the data shown in Fig. 3 at 1680 psi and 100 psi. That the sum of the probable errors of the periods is only about one-third the total change in period due to pressure further confirms the pressure effect. The change in phase is about twice the sum of the probable errors in phase. The value of the period at one atmosphere, 1.57×10^{-5} gauss^{-1}, is in good agreement with the value 1.5×10^{-5} gauss^{-1} observed by Reynolds *et al.*[8] The theories[9–11] provide for a phase of $\pi/4$ while we find experimentally the value $1.5\times\pi/4$ for the phase of the Hall coefficient oscillations. We have no explanation as to why there should be a phase change due to pressure.

II. Amplitude and Monotonic Part

In view of the lack of a theoretical expression for the amplitude of the Hall coefficient oscillations, it is of interest to ascertain an empirical form for the observed dependence of amplitude on magnetic field. It was found by a rather careful numerical analysis[18] that an expression of the general form,

$$a_y=\frac{c_1}{H^{c_2}\sinh(c_3/H)}, \tag{2}$$

is a suitable representation of the observed amplitude within the experimental error. A quite good fit was obtained with

$$c_2=2.4\pm0.1,$$
$$c_3=(67\pm1)\times10^3 \text{ gauss}.$$

The de Haas-van Alphen susceptibility oscillations are described theoretically by the same general form,

TABLE I. Effect of pressure on the period and phase of Hall coefficient oscillations in bismuth at 4.2°K and the corresponding pressure coefficients (assuming linear dependence of period and phase on pressure). The probable errors in the period and phase are given.

Pressure	Period, β/E_0 (gauss^{-1})	Phase, θ
100 psi (6.8 atmos)	$(1.57^2\pm0.003)\times10^{-5}$	$(0.73^4\pm0.04^5)\times\pi/2$
1680 psi (114 atmos)	$(1.59^2\pm0.004)\times10^{-5}$	$(0.88^8\pm0.04^4)\times\pi/2$

Pressure coefficients
$\Delta(\beta/E_0)/\Delta p=1.9\times10^{-13}$ gauss^{-1}-dyne^{-1}-cm^2
$\Delta\theta/\Delta p=1.4\times10^{-9}\times\pi/2$ dyne^{-1}-cm^2

[17] E. T. Whittaker and G. Robinson, *The Calculus of Observations* (Blackie and Sons, Ltd., London, 1944).

[18] Equation (2) was first solved, using three experimental values of a_y, for preliminary values, c_1', c_2', and c_3'. Treating c_2 as a parameter in the neighborhood of c_2' we let $z=\log_{10}\{a_y\sinh(c_3/H)\}$, $x=\log_{10}(H^{-1})$, so that $z=c_2x+\log_{10}(c_1)$, i.e., the equation of a straight line. For each c_2 value assumed as a parameter, c_2 and $\log_{10}(c_1)$ of the straight line function were determined by least squares using six experimental values of a_y and the corresponding H^{-1} values for fields ranging from 6000 to 12 000 gauss. The sum of the squares of the residuals of z from the least-squares-fitted line were then plotted as a function of the assumed parameter, c_2. This curve had a definite minimum at a c_2 value of $(67\pm1)\times10^3$. Such a procedure gives perhaps the best possible fit of the experimental amplitude by the formula Eq. (2).

but with

$$c_2=\tfrac{3}{2}, \tag{3a}$$

$$c_3=2\pi^2kT/\beta. \tag{3b}$$

The numerical value, $c_3=25\times10^3$ gauss, has been determined from Eq. (3b) using a value of β calculated from Shoenberg's E_0 value[12] (which he computed from susceptibility data) and our observed period, β/E_0, and $T=4.21$°K. In view of the wide discrepancy between these c_3 values, the use of Eq. (3b) for describing the Hall coefficient c_3 is not justified.

An analysis of the field dependence of the amplitude at 100 psi and at 1680 psi has shown that there is no change, within experimental error, of the amplitude function of the Hall coefficient oscillations due to pressure in the pressure range up to 114 atmos. The experimental results were accurate enough to permit a reliable test of the form of Eq. (2) but were not good enough to permit a determination of a change of c_2 or c_3 due to pressure. This fact, together with the fact that Eq. (3b) is not a suitable form for representing c_3 for the Hall coefficient amplitude, has ruled out the possibility of ascertaining the value of β and a change in β, if any, resulting from pressure. Thus any quantitative discussion of the detailed change in the overlap configuration due to pressure is not now possible.

The observed monotonic part representing f_y in Eq. (1) may also be fitted quite well by Eq. (2) with the proper set of parameters. However, the value of c_3 suitable for representing f_y by Eq. (2) turns out to be about one-tenth the value of c_3 needed in representing a_y. If Eq. (2) is ultimately to be the appropriate theoretical form for both a_y and f_y, then it might be expected that at least the value of c_3 should be common to both. Thus it is felt that Eq. (2) is not the correct description for the monotonic part even though it may be proper, at least in form, for the amplitude. In view of the lack of a suitable theoretical representation for the monotonic part of the Hall coefficient, it is perhaps more appropriate to give the actual observed values at high and low pressures. At 6.8 atmos some values are, in cm³/coulomb, −13 262, −11 002, and −9225 respectively at reciprocal field values of 8729, 10 322, and 11 900 units of 10^{-5} gauss. When the pressure was increased to 114 atmos the corresponding changes were to the values, in cm³/coulomb, −13 350, −11 082, and −9297. Thus, for example, at 11 400 gauss the monotonic part of the Hall coefficient increased in absolute value by about one percent due to the pressure increase of 107 atmos. This increase is much larger than the experimental errors involved.

SUMMARY

We have observed a pressure dependence of the period, phase, and the monotonic part of the Hall coefficient variation with magnetic field in a bismuth crystal oriented with the trigonal axis parallel to the magnetic field direction. The period was found to be in agreement with the period of the Hall coefficient oscillations observed by Reynolds *et al.* for the same orientation. The period was found to increase by 1.3 percent due to changing the pressure from 6.8 atmos to 114 atmos. The change in phase due to this pressure change is about 20 percent. The monotonic part of the Hall coefficient was found to increase in absolute value, changing, for example, from −13 262 cm³/coulomb at 11 400 gauss and 6.8 atmos to −13 350 cm³/coulomb at 11 400 gauss and 114 atmos. There was no appreciable change in the amplitude of the oscillations because of the 107-atmos change in pressure.

Our experimental results have been considered in light of the theory for the de Haas-van Alphen susceptibility oscillations. It does not appear that the theoretical expression for the amplitude of the susceptibility can be extended to the amplitude for the Hall coefficient oscillations. However, Eq. (2) is empirically suitable for representing the amplitude for the particular orientation studied.

The change in period is explained qualitatively if it is assumed that, due to the application of pressure, the Brillouin zone boundary moves a different amount in **k**-space than the highest occupied Fermi level resulting in a net change in the depth of overlap of energy surfaces. We have no explanation for the change in phase due to pressure.

Measurements of the magnetic susceptibility oscillations in bismuth over a temperature range from 4.2°K up to a temperature at which they are no longer observable would be of interest. A temperature-dependent period might be expected as in the case of zinc.[19] Such data along with the pressure-dependent data reported above might permit an assessment of the relative importance of the effects upon period, and hence overlap, due to changes in lattice dimensions as compared with the electron transfer mechanism suggested by Elcock[20] in explaining the temperature dependence of field-independent susceptibilities. The pursuance of such experimental studies might ultimately lead to the assignment of those effects due separately to overlapping electrons and underlapping holes. At present, it would be unsafe to draw conclusions in regard to the temperature dependence of the overlap in bismuth through correlations of its pressure-dependent properties at 4.2°K and the temperature-dependent properties of zinc at atmospheric pressure.

[19] T. G. Berlincourt and M. C. Steele, Phys. Rev. **95**, 1421 (1954).

[20] E. W. Elcock, Proc. Roy. Soc. (London) **A222**, 239 (1954).

Fast Neutron Bombardment of p-Type Germanium

J. W. Cleland, J. H. Crawford, Jr., and J. C. Pigg
Oak Ridge National Laboratory, Oak Ridge, Tennessee

(Received May 11, 1955)

Extensive studies of fast neutron bombardment effects in p-type Ge have been carried out. Analysis of the initial conductivity behavior indicates the introduction of three vacant states below the middle of the forbidden band: (1) an acceptor state 0.18 ev above the valence band, (2) an occupied (donor) state which acts as a hole trap with a depth of ~0.07 ev, (3) a second low-lying donor which is a quite shallow hole trap. This distribution of energy levels is consistent with the Frenkel defect model of James and Lark-Horovitz. After prolonged bombardment or room temperature aging, the ionization energy of bombardment-produced acceptors moves to lower values, a behavior which has been attributed to a difference in annealing rate of interstitials and vacancies. The scattering mechanism for holes in bombarded specimens does not appear to be completely describable in terms of ionized impurity scattering. Moreover, the temperature dependence of hole mobility is quite different from electron mobility in bombarded specimens. Extensive bombardments in the range from −165° to −90°C are described and complex relaxation effects observed on warming are discussed.

I. INTRODUCTION

THE effect of fast neutron bombardment of the semiconducting properties of n-type Ge was treated in some detail in a previous paper.[1] The behavior of electron concentration and mobility during exposure in the reactor at or near room temperature was described and studies of conductivity and photoconductivity during and after exposure at low temperatures were also discussed. It was shown that the localized states produced by bombardment in the room temperature range, which are responsible for the removal of conduction electrons, are consistent with the energy level model for Frenkel defects proposed by James and Lark-Horovitz.[2] It was further shown that the effect of irradiation on Hall mobility could not be explained simply on the basis of additional charge-center scattering and that low-temperature irradiation was accompanied by extensive photoconductivity which could be interpreted as arising from trapped minority carriers.

The present paper is concerned with fast-neutron bombarded p-type Ge: both originally p-type material and that converted to p-type by irradiation. The situation with regard to changes in carrier concentration is not so clear cut as in the case of n-type material because of the greater complexity of defect-produced energy levels below the middle of the forbidden band. Theory predicts[2] and experiment confirms[3] the presence of both occupied and vacant localized states above the valence band. Consequently, depending on the initial hole concentration, the conductivity may either increase or decrease with bombardment.

Extensive measurements of conductivity σ on specimens with a wide range of initial hole concentration p_0 during irradiation in both the room-temperature range and at −78°C have been obtained. These data, together with measurements of Hall coefficient and conductivity as a function of temperature after successive periods of exposure at room temperature, permit an extensive examination of the nature and position of the states associated with defects which lie below the middle of the forbidden band. Information concerning hole mobility in irradiated Ge is also presented. Finally, the results of low-temperature exposures in the range from −165°C to −90°C are discussed.

II. IMPLICATIONS OF THE MODEL OF JAMES AND LARK-HOROVITZ

It was shown previously[3] that both the magnitude and the sign of the initial rate of change of hole concentration $[dp/d(nvt)_f]_{i=0}$ are dependent on the initial hole concentration p_0 and the temperature of exposure T_e. This result is consistent with the proposed model of energy levels associated with Frenkel defects,[2] since both occupied and vacant localized states above the top of the valence band are predicted by this model. Consequently, the situation is characterized by two competing processes: production of holes by the vacant states (acceptors) and trapping of holes by the occupied states (deep-lying donors). Under conditions of thermal equilibrium, the effect of Frenkel defect introduction depends on the position of the initial Fermi level ζ^0 relative to the limiting value ζ^* which is determined solely by the positions of the defect energy levels and the temperature. According to the model, two occupied levels and one vacant level are expected. If it is assumed that all of the chemical acceptors are completely ionized, the law of electrical neutrality requires that

$$p = p_0 + N_1^- - N_2^+ - N_3^+, \tag{1}$$

where N_1^- is the concentration of ionized acceptors and N_2^+ and N_3^+ are the concentration of holes in the deep and shallow hole traps, respectively. If it is assumed that the levels are discrete, the law of mass

[1] Cleland, Crawford, and Pigg, Phys. Rev. **98**, 1742 (1955).
[2] H. M. James and K. Lark-Horovitz, Z. Physik. Chem. **198**, 107 (1951).
[3] Cleland, Crawford, Lark-Horovitz, Pigg, and Young, Phys. Rev. **84**, 861 (1951).

action gives

$$N_1^- = NK_1/(p+K_1),$$
$$N_2^+ = Np/(p+K_2), \quad (2)$$
$$N_3^+ = Np/(p+K_3),$$

where $N_1 = N_2 = N_3 = N$, the total Frenkel defect concentration, and the equilibrium constants are given by

$$K_i = \gamma_i C(T) e^{-\epsilon_i/kT}, \quad (3)$$

where

$$C(T) = (2\pi m_h kT/h^2)^{\frac{3}{2}},$$

ϵ_i is the energy of the level measured from the top of the valence band, and γ_i takes a value of either unity or four depending on whether the state occupied by the hole has a residual electronic spin or not.[4] Substitution of Eqs. (2) in Eq. (1) yields

$$p = p_0 + N\left[\frac{K_1}{p+K_1} - \frac{p}{p+K_2} - \frac{p}{p+K_3}\right], \quad (4)$$

a quartic equation in p. If

$$p_0 = p^* = 2C(T)\exp(-\zeta^*/kT), \quad (5)$$

where p^* is the limiting hole concentration for the temperature in question, p will remain invariant with defect introduction. Under these conditions Eq. (5) reduces to the cubic equation

$$p^{*3} + \tfrac{1}{2}(K_1+K_2+K_3)p^{*2} - \tfrac{1}{2}K_1K_2K_3 = 0. \quad (6)$$

Only the positive root of Eq. (6) is significant, which, to a good approximation, is given by

$$p^* = \left(\frac{K_1K_2K_3}{K_1+K_2+K_3}\right)^{\frac{1}{2}}. \quad (7)$$

Using Eqs. (5) and (7), Eq. (3) becomes

$$\zeta^* = \frac{\epsilon_1+\epsilon_2+\epsilon_3}{2} - \tfrac{1}{2}kT\ln\left[\frac{4(\gamma_1 e^{-\epsilon_1/kT} + \gamma_2 e^{-\epsilon_2/kT} + \gamma_3 e^{-\epsilon_3/kT})}{\gamma_1\gamma_2\gamma_3}\right]. \quad (8)$$

If ϵ_3 is significantly smaller than ϵ_2 and hence ϵ_1, Eq. (8) may be approximated by

$$\zeta^* = \tfrac{1}{2}(\epsilon_1+\epsilon_2) - \tfrac{1}{2}kT\ln(4/\gamma_1\gamma_2), \quad (9)$$

which is identical to the expression obtained for a single trapping level and a single acceptor. The temperature dependence of ζ^* enters through the logarithmic term of Eq. (9) which, in this approximation, is

$$d\zeta^*/dT = -\tfrac{1}{2}k\ln(4/\gamma_1\gamma_2). \quad (10)$$

From the nature of the levels below the middle of the band as suggested by the theory,[2] it can be shown, whatever the identification of levels in question with those available, that $\gamma_1\gamma_2 = 4$. Hence $d\zeta^*/dT = 0$.

The value of ζ^* indicates only the midpoint between the acceptor and the deepest hole trap. It is possible in principle to determine ϵ_1 and ϵ_2 from initial rates of change of p during bombardment as a function of p_0, provided that the relation between the Frenkel defect concentration N and integrated flux is known. The derivative of Eq. (4) with respect to N evaluated at $N=0$ and $p=p_0$ gives for the initial rate of change

$$[dp/dN]_{N=0} = \frac{K_1}{p_0+K_1} - \frac{p_0}{p_0+K_2} - \frac{p_0}{p_0+K_3}. \quad (11)$$

Hence there are apparently three adjustable parameters which would seem to indicate that curve fitting with Eq. (11) may not be a sensitive test of the theory. In actual fact, however, the condition imposed by either Eq. (8) or Eq. (9) reduces the number of parameters to two. Furthermore, the contribution of the last term of Eq. (11) is negligible until ζ^0 lies within $\sim 2kT$ of ϵ_3. In order to test the theory rigorously, it is necessary to use the values of ϵ_1, ϵ_2, and ϵ_3 obtained by using Eq. (11) to reproduce a representative conductivity *vs* bombardment curve according to Eq. (4). The latter test is possible only if (1) the effect of bombardment on mobility is negligible compared to the effect on hole concentration, (2) there is no differential annealing of different types of defects during exposure and (3) appreciable photoconductivity is absent. In the following section we shall apply these relationships to the behavior of conductivity measured during bombardment at room and dry ice temperatures.

III. EFFECT OF BOMBARDMENT ON CONDUCTIVITY

A. Initial Rate of Change

The conductivity σ of a large number of single-crystal p-type samples with a wide range of hole concentrations has been measured during bombardment in the reactor in the room-temperature range. Under the assumption that the initial change of mobility is negligible compared to the effect on hole concentration, the initial change in hole concentration per incident fast neutron[5] $[dp/d(nvt)_f]_{i=0}$ was calculated from σ *vs* nvt curves, using the initial conductivity σ_0 and uncorrected values[6] of p_0 obtained from Hall coefficient measure-

[4] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, London, 1948), second edition, pp. 157–158. See also J. H. Crawford, Jr., and D. K. Holmes, Proc. Phys. Soc. (London) **A67**, 294 (1954). The necessity of using this statistical factor was pointed out to the authors by H. M. James.

[5] The flux unit is discussed in reference 1.

[6] Here, as elsewhere in the paper, hole concentrations were calculated from the customary relation $p = 7.37\times10^{18}/R$, where R is the Hall coefficient. No attempt was made to correct the values for impurity scattering since recent work [F. J. Morin, Phys. Rev. **93**, 62 (1954)] has shown that the ratio of Hall to drift mobility μ_H/μ_d for holes not only does not have the proper magnitude but is marked by temperature dependence. Willardson, Harman, and Beer [Phys. Rev. **96**, 1512 (1954)] have shown that this discrepancy is produced by the complex nature of the valence band in Ge. Consequently, values of p given in this paper may be in error by as much as a factor of two.

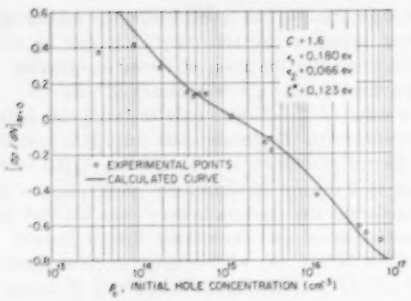


FIG. 1. The initial rate of change of hole concentration per Frenkel defect *vs* the initial hole concentration during bombardment at 195°K. The points are experimental and the solid curve is calculated from Eq. (11) assuming that C, the number of defects introduced per incident neutron, is 1.6.

ments. Values of p_0 and the exposure temperature for these specimens are listed in Table I. Although there is scatter in the data, it is apparent that $[dp/d(nvt)_f]_{t=0}$ changes from positive to negative with increasing p_0 at $p^*=(6.5\pm1.5)\times10^{16}$ cm^{-3}. The value of ζ^* at this temperature ($\sim$50°C) may be obtained from Eq. (5) provided m_h is known. Measurements of magnetic susceptibility[7] and cyclotron resonance[8] indicate that $m_h\simeq-0.3m_0$. With this value of m_h, $\zeta^*(323°K)=0.120\pm0.006$ ev. Besides room-temperature exposures, a number of specimens have been irradiated at dry ice temperature. At −78°C, $p^*=(1.5\pm0.2)\times10^{15}$ cm^{-3} leading to $\zeta^*=0.123\pm0.003$ ev. Hence it appears that ζ^* is temperature independent within experimental error. This would indicate that the approximation leading to Eq. (9) is a good one.

Attempts have been made to evaluate ϵ_1, ϵ_2, and ϵ_3 by using Eq. (11) to fit the $[dp/d(nvt)_f]_{t=0}$ *vs* p_0 curve for the −78°C exposures. It has been shown[9] that $\sim$3.2 electrons are removed per incident fast neutron in moderate to high-conductivity n-type material. Since two electron traps are expected per Frenkel defect,[2] this result indicates that the number of defects per incident neutron is approximately 1.6. The results of this analysis are shown in Fig. 1 where $[dp/dN]_{N=0}$ is plotted against $\log p_0$. The points are experimental and the solid curve was calculated from Eq. (11) by using $\epsilon_1=0.180$ ev, $\epsilon_2=0.066$ ev, and $\epsilon_3\simeq0$ and assuming $\gamma_1=4$ and $\gamma_2=1$. Considering expected experimental uncertainties, a reasonable fit to the data was obtained. Interchange of the values of γ_1 and γ_2 did not permit a good correlation even for a wide range of choice of energy parameters. These results strongly indicate that, provided the model is valid, ϵ_3 is quite small and, therefore, only one hole trap need be considered at this and higher temperatures. Furthermore, the required choice of values for γ_1 and γ_2 indicates that the acceptor can be identified with the doubly ionized interstitial atom. This calculation assumes that there has been no appreciable differential annealing between interstitials and vacancies in the initial range.

B. Conductivity *vs* Bombardment Curve

The energy parameters obtained above were used in Eq. (4) to calculate σ *vs* $(nvt)_f$ curves obtained at −78°C under the assumption that the mobility is unaffected by exposure. Although reasonable correspondence was obtained early in the exposure, the calculated and observed curves began to diverge as the bombardment proceeded. It is significant that the calculated curve predicts a lower value of σ after appreciable exposure than is actually observed for a p_0 either greater or less than p^*. If a decrease in mobility during exposure was solely responsible for the discrepancy, a deviation in the opposite direction would be expected. The greater observed values of σ may be due to either or both of two processes: (1) trapped minority carriers which produce an excess concentration of majority carriers, enhancing σ; and (2) a difference in annealing rate of the two types of defects which would alter the relative concentration of energy levels arising from each type and, thereby the distribution of energy levels. If the second process is responsible, the alteration of the distribution must occur in such a direction to cause p^* apparently to increase as bombardment progresses. Evidence for both of these processes will be discussed below in more detail.

TABLE I. Initial rate of change of hole concentration during pile ambient temperature bombardment.

Sample	p_0 (cm^{-3})	T,°C	$[dp/d(nvt)_f]_{t=0}$
1	4.2×10^{14}	20	0.70
2	1.7×10^{14}	30	0.77
3	3.5×10^{15}	53	0.08
4	4.0×10^{16}	55	0.20
5	5.7×10^{16}	62	0.037
6	7.2×10^{16}	49	−0.058
7	8.1×10^{16}	50	−0.10
8	1.5×10^{17}	62	−0.05
9	5.3×10^{17}	45	−0.80
10	7.0×10^{17}	32	−0.50
11	1.3×10^{19}	48	−2.2
12	1.5×10^{19}	48	−2.9

The results of the foregoing analysis of the initial behavior of conductivity during bombardment reveals that the model proposed by James and Lark-Horovitz is consistent with observation. When applied to the behavior on prolonged bombardment, qualitative agreement is still obtained, but the data indicate that additional processes appear which require explanation. In this connection, it should be pointed out that similar

[7] D. K. Stevens and J. H. Crawford, Jr., Phys. Rev. **92**, 1065 (1953).
[8] Dresselhaus, Kip, and Kittel, Phys. Rev. **92**, 827 (1953); Dexter, Zeiger, Lax, and Rosenblum, Phys. Rev. **95**, 597 (1954).
[9] Cleland, Crawford, Lark-Horovitz, Pigg, and Young, Phys. Rev. **83**, 312 (1951).

studies carried out with 10-Mev deuterons[10] at a much higher rate of defect introduction show some interesting differences. The value of ζ^* obtained by fitting a σ *vs* bombardment curve at -78°C was 0.089 ev, considerably smaller than the value obtained above from initial behavior. In addition, specimens with p_0 only slightly greater than the value of p^* obtained with fast neutrons, showed first a decrease in σ followed by an increase after additional deuteron bombardment. Although it is possible that the different results may be produced by differences in the effect of fast neutrons and charged particles, it is interesting to note that, considering the much higher rate of disordering by deuterons, these deviations are in the direction expected if process (2) described previously were operating.

IV. TEMPERATURE DEPENDENCE OF HOLE CONCENTRATION AFTER EXPOSURE

In general, the hole concentration before exposure is only slightly temperature dependent. Since, according to the model in question, ζ^* is essentially temperature independent and p^* is given by Eq. (5), p^* is expected to correspond to p_0 at some temperature T^* in an easily attainable temperature range (77°K to 300°K) over a wide range of p_0 values (10^{13} to 10^{17} cm^{-3}). Bombardment should cause an increase in p above T^* and a decrease in p below T^*. This result should be reflected in a rotation of the log$R$ *vs* $1/T$ curve about a point $1/T^*$ approaching a slope of ζ^*/k.

Curves have been obtained after various exposures on a number of p-type specimens. Because of the induced radioactivity in the specimen and holder, it was usually necessary to wait appreciable periods (1 to 3 days) before handling was safe. Since room-temperature aging may occur during this waiting period, the curves do not necessarily represent conditions existing immediately on removal from the reactor. Typical log R *vs* $1/T$ curves after bombardment are compared with their original curves in Fig. 2 for three widely different values of p_0. The dashed curve in Fig. 2(a) was taken 3.5 days after an exposure to 1.6×10^{-17} fast neutrons cm^{-2} at 60°C on a specimen with $p_0=3.4\times10^{13}$ cm^{-3}, that in Fig. 2(b) was obtained 6 days after $(nvt)_f = 5.8\times10^{16}$ cm^{-2} at 60°C with $p_0=7.0\times10^{16}$ cm^{-3}, and that in Fig. 2(c) 19 hr after $(nvt)_f=2.2\times10^{16}$ cm^{-2} at 60°C with $p_0=7.3\times10^{14}$ cm^{-3}. The solid curves are those taken before exposure.

The curve in Fig. 2(a) indicates that bombardment decreases the hole concentration over the whole temperature range. This behavior is borne out by the observed conductivity change during exposure. Further irradiation was found to produce an additional decrease in σ toward a limiting value. In the case of the curve in Fig. 2(b), the curve after exposure crosses the original curve indicating an increase in p above 250°K and a decrease in p at lower temperatures. The behavior in these two cases is qualitatively what is expected according to the model employed. In the case of small p_0, however, indications of Fig. 2(c) are that there is an increase in p over the whole temperature range despite the fact that the value of ζ^* indicates that the two curves should cross at $T^*=166$°K. Furthermore, the curvature after bombardment suggests a wide distribution of acceptor level energies that vary from an appreciable value at room temperature to a value near zero at liquid nitrogen temperature. These results indicate that, although there is qualitative agreement with the model at high values of p, there is a serious lack of agreement for small hole concentrations. The discrepancy cannot be reasonably explained by the initial introduction of an appreciable concentration of very shallow acceptors during bombardment since they would, to a large extent, mask the presence of hole traps indicated by studies of initial behavior (Sec. III A). Rather it is necessary to assume that the apparent low acceptor ionization energy results from some process which occurs during bombardment or during the aging period at room temperature. The same two processes appear probable which were mentioned in Sec. III B, namely, trapping of minority carriers or differential annealing of interstitials and vacancies. Evidence for these two processes will be treated in turn.

Although a more detailed discussion of minority carrier trapping will be deferred until later (Sec. VI), it is appropriate to say here than on closer examination

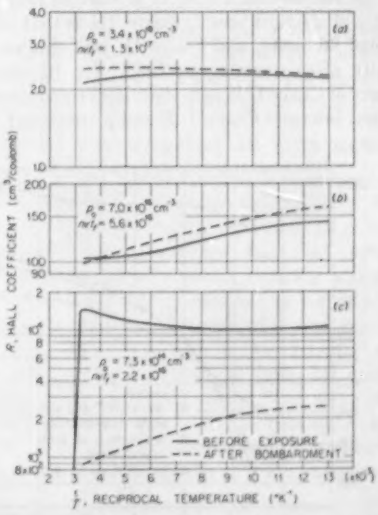


FIG. 2. Log Hall coefficient *vs* reciprocal temperature for three p-type specimens. The solid curves refer to the original behavior and the dashed curve was taken after exposure. The exposures $(nvt)_f$ and initial hole concentrations are indicated on the curves.

[10] Forster, Fan, and Lark-Horovitz, Purdue Progress Reports to Signal Corps, PRF-746, January 1, 1952 (unpublished), pp. 28–31. See also J. H. Forster, thesis, Purdue University, 1953 (unpublished).

it was found that at liquid nitrogen temperature, the Hall coefficient and resistivity of irradiated samples were extremely photosensitive. All of the data, however, were taken on specimens wrapped with several layers of linen tape and mounted in a narrow-necked Dewar, the mouth of which was plugged with glass wool. Consequently, the amount of ambient light incident on the specimen was quite small. It was impossible, however, to exclude the β and γ radiation which resulted from radioactive decay of the specimen, leads and holder. This radiation, though of relatively low intensity, is capable of producing electron-hole pairs. Consequently, measurements at low temperature taken shortly after removal from the pile undoubtedly include photoeffects. This process, however, cannot be solely responsible for the discrepancy, since after the activity in the specimen and its surroundings had essentially completely decayed, the difference between the observed and expected curve persisted.

The possible alteration of the distribution of energy levels was examined by following the progress of aging at and slightly above room temperature. The two specimens which were studied both showed a flattening of the log R *vs* $1/T$ curve while standing at room temperature. In one of these, the curve after exposure crossed the original one and it was difficult to decide whether the flattening of the curve was a result of uniform annealing of the vacancies or an alteration of the energy level distribution. The second specimen, however, gave a curve similar to that of Fig. 2(c), and its interpretation is less ambiguous. The results of the study are shown in Fig. 3. Curve I is the log R *vs* $1/T$ curve before bombardment, Curve II is that 10 hours after an $(nvt)_f$ of 2.6×10^{16} cm^{-2}, Curve III is that after an additional 96 hours, and Curve IV was obtained subsequently after warming to 140°C for 10 min. The curvature of Curve II is such that, if its trend continues, it will not intersect Curve I. R and ρ, measured during subsequent aging, showed a slow increase at room temperature and a slow decrease at 77°K. After 106 hours total aging the rate of change was very small and Curve III resulted. If both interstitials and vacancies were removed at the same rate, the curves after bombardment would be expected to approach the original one rather uniformly over the whole range. Comparison of Curves II and III, however, show that this is not the case. Instead, the decrease at low temperature and the increase at high temperature indicates that the distribution of acceptors is shifted toward lower ionization energy. In addition, since a rotation, instead of a general lowering, of Curve III occurs with respect to Curve II, the total concentration of acceptors apparently decreases during aging. On heating the specimen at 140°C for 10 min, it is evident from Curve IV that the process is accelerated and marked flattening of the curve occurs.

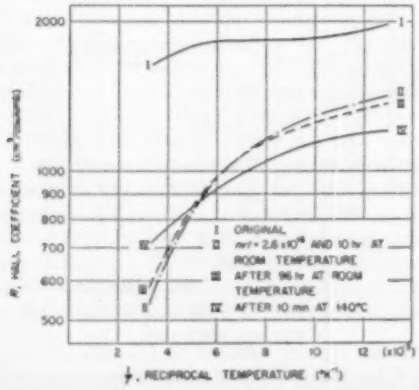


FIG. 3. Log Hall coefficient *vs* reciprocal temperature for a p-type specimen before and after bombardment and subsequent to annealing at room temperature and 140°C.

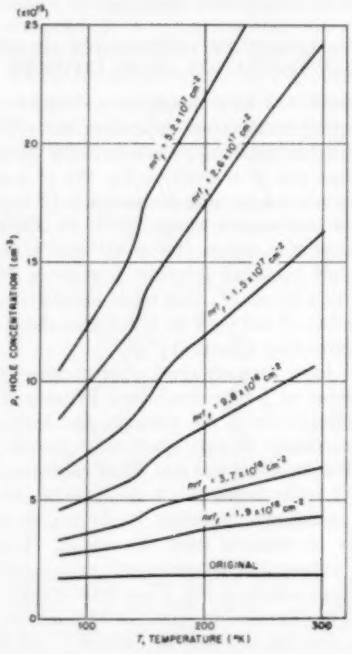


FIG. 4. Hole concentration *vs* temperature after successive exposures. The exposures are indicated on the curves.

In view of these observations it is not unreasonable to assume that similar annealing occurs during exposure as well. According to the model in question and indications of Sec. III A, a vacancy introduces two low-lying empty states which are filled by electrons from a doubly ionized interstitial. If an interstitial is removed by some process, such as clustering or trapping in the vicinity of a dislocation, two shallow acceptors are produced. Since a decrease in total acceptor concentra-

tion is also observed, recombination, to a smaller extent, of interstitials and vacancies must also be postulated. It is noteworthy that these assumptions require the interstitial atom to be the mobile component of the Frenkel defect. These considerations receive support from observations on low-temperature aging of irradiated n-type Ge[1] and quenching experiments of Mayburg.[11]

The temperature dependence of hole concentration[6] after successive periods of bombardment at ~40°C was determined from Hall coefficient measurements on one p-type specimen ($p_0 = 1.8\times10^{15}$ cm^{-3}). Representative p *vs* T curves after various exposures are shown in Fig. 4. All of the curves after bombardment are characterized by a sigmoid involution, with its inflection point centered in the vicinity of 145°K, superimposed on an almost linear increase in p with T. The almost linear increase of p indicates a broad distribution of acceptor ionization energies, whereas the sigmoid step reveals the presence of a discrete level or closely grouped set of levels superimposed on the broad distribution. Although it is possible to obtain acceptable fits to these curves by using arbitrarily chosen acceptor distributions,[12] such fits require the use of at least four adjustable parameters. Consequently, little can be learned concerning the nature of the broad distribution by these means. Such calculations, however, do indicate that the discrete level or bump in the distribution lies in the vicinity of 0.10 ev above the band edge.[12]

In the case of an n-type specimen converted to p-type by bombardment, the resulting energy-level distribution may be examined without complications produced by holes initially present. Log R *vs* $1/T$ curves obtained after bombardment on an originally n-type specimen ($n_0 = 7.5\times10^{15}$ cm^{-3}) are shown in Fig. 5. Curve I was obtained after an $(nvt)_f = 1.0\times10^{16}$ cm^{-2}, Curve II after an $(nvt)_f = 2.2\times10^{16}$ cm^{-2}, Curve III after a 40-day anneal at room temperature, and Curve IV after annealing for 30 min at 150°C. Shortly after conversion to p-type, the slope of the log R *vs* $1/T$ curve (Curve I) is linear over most of the range with a slope corresponding to an acceptor ionization energy of ~0.08 ev. This level probably corresponds to the one producing the step in Fig. 4 and may presumably be identified with the deep trap determined from dry ice experiments (see Sec. III A). The curvature toward lower slope near 77°K indicates the presence of more shallow levels. After additional irradiation (Curve II) the shift of acceptor ionization toward lower values is quite

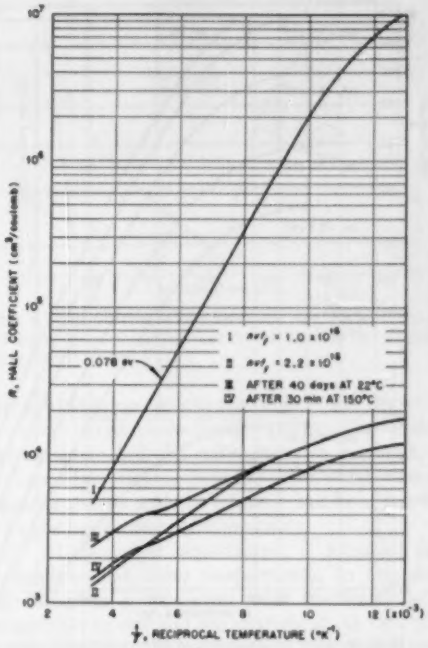

R, HALL COEFFICIENT (cm³/coulomb)
1 nvt_f = 1.0 x10¹⁶
2 nvt_f = 2.2 x10¹⁶
3 AFTER 40 days AT 22°C
4 AFTER 30 min AT 150°C
0.078 ev
1/T, RECIPROCAL TEMPERATURE (°K⁻¹)
(x10⁻³)

FIG. 5. Log Hall coefficient *vs* reciprocal temperature for an originally n-type specimen converted to p-type by bombardment, after two successive bombardments and subsequent anneals.

evident. After the 40-day room-temperature anneal (Curve III), behavior is observed which is similar to that occurring in the case of initially p-type specimens (Fig. 3). After 40 days, however, the room-temperature aging process is essentially complete as is evidenced by the uniform decrease of p over the whole range which accompanies the 150°C annealing period (Curve IV).

V. EFFECT OF BOMBARDMENT ON HOLE MOBILITY

Studies of n-type Ge[1] reveal that fast-neutron bombardment markedly decreases electron mobility. Moreover, the additional scattering was found to possess a greater temperature dependence than that predicted by the usual theories of ionized impurity scattering. Similar studies have been carried out on p-type Ge.

In Fig. 6 the mobility μ[13] is plotted against temperature for the specimen of Fig. 4 ($p_0 = 1.8\times10^{15}$ cm^{-3}) after several exposures. Values of $(nvt)_f$ for each exposure are indicated on the curves. The slope of the curve before irradiation is -2.1, in reasonable agreement with the reported temperature dependence for lattice scattering in p-type Ge.[14,15] Furthermore, the

[11] S. Mayburg, Phys. Rev. **95**, 38 (1954).

[12] The authors have employed with good results a uniform distribution extending from the edge to the center of the forbidden band with a superimposed discrete level centered ~0.10 ev above the band edge. In addition I. M. Katz (unpublished work) has obtained an excellent fit using a Gaussian distribution of levels of 0.24-ev half-width centered at the middle of the forbidden band with a superimposed discrete level at ~0.09 ev. These energy values were calculated assuming $m_h = m_0$; hence they are expected to be somewhat high.

[13] As used here, $\mu = R\sigma/r$, where the Hall parameter r is taken to be $3\pi/8$, the same value as that used in deriving ρ. Although the true value of r may be quite different from this (see reference 6), it is desirable to use the same value throughout, since r is thus effectively eliminated when $1/\mu$ is compared with ρ (Fig. 7).

[14] M. B. Prince, Phys. Rev. **92**, 681 (1953).

[15] F. J. Morin, Phys. Rev. **93**, 62 (1954).

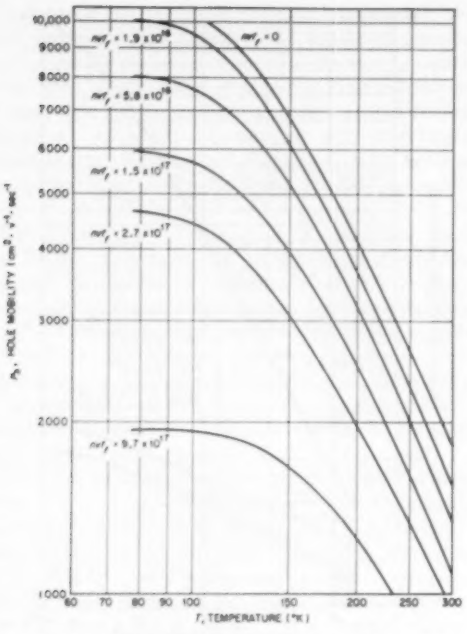

$nvt_f = 0$
$nvt_f = 1.9 \times 10^{16}$
$nvt_f = 5.8 \times 10^{16}$
$nvt_f = 1.5 \times 10^{17}$
$nvt_f = 2.7 \times 10^{17}$
$nvt_f = 9.7 \times 10^{17}$
μ_p, HOLE MOBILITY (cm², v⁻¹, sec⁻¹
T, TEMPERATURE (°K)

FIG. 6. Hole mobility *vs* temperature after successive exposures. The exposures are indicated on the curves.

initial curve shows the expected onset of the ionized impurity scattering at low temperatures, which is not sufficiently great at this concentration to produce a maximum in the curve in this temperature range (77° to 300°K). On bombardment μ decreases over the whole temperature range with a somewhat greater decrease occurring at low temperature. It is interesting to note that, even after appreciable bombardment, no discernable maximum is produced. If the additional, bombardment induced scattering were strictly of the charge-center type, a maximum which moves toward higher temperature with increasing defect introduction might be expected to appear. The lack of such a maximum may be taken as tentative evidence that the scattering associated with lattice damage of this type may not be amenable to treatment in terms of ionized impurity scattering.[16–18] It is noteworthy, however, that the excessively large temperature dependence of electron mobility observed in bombarded *n*-type Ge[1] is absent in *p*-type. Similar behavior to that shown in Fig. 6 was observed with a number of other specimens. In no case was a maximum observed which was not already present before exposure.

Since the curves of Fig. 6 indicate that μ at 77°K is limited predominantly by scattering processes other than lattice scattering, the reciprocal of this value $1/\mu_{77}$ is expected to be approximately proportional to the scattering probability of these processes at that temperature. The dependence of the scattering probability on defect concentration is shown in Fig. 7 where $1/\mu_{77}$ is plotted against the hole concentration at 77°K p_{77}. The convenient use of p_{77} as an index of disorder is possible because it increases approximately linearly with exposure. All points after the initial exposure fall on a straight line whose slope is 2.1×10^{20} in appropriate units. Since the same value of the Hall parameter was used in determining both μ and p, the correlation should be independent of its value.[13]

Although it was indicated in the foregoing that the scattering mechanism of the defects may not be of the ionized-impurity type, it can be shown that the data of Fig. 7 are not completely inconsistent with such a model. According to the Frenkel-defect energy-level model,[2] it is reasonable to expect the concentration of ionized scattering centers to be considerably larger than p_{77}. Moreover the model suggests that an appreciable fraction of these may be doubly rather than singly charged. For a mixture of scattering charges it is necessary to write the Conwell-Weisskopf formula[16] in a more general form.[19] In terms of the measure of

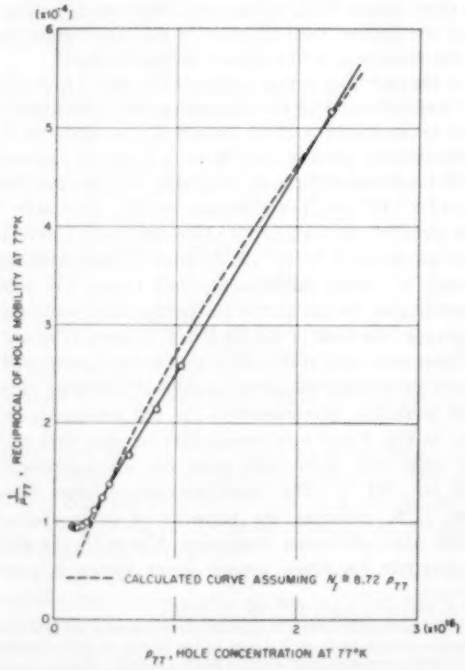

(x10⁻⁴)
$\frac{1}{\mu_{77}}$, RECIPROCAL OF HOLE MOBILITY AT 77°K
CALCULATED CURVE ASSUMING $N_i = 8.72\ p_{77}$
p_{77}, HOLE CONCENTRATION AT 77°K
3 (x10¹⁶)

FIG. 7. The reciprocal of hole mobility *vs* hole concentration at 77°K in a *p*-type specimen after successive bombardments.

[16] E. Conwell and V. F. Weisskopf, Phys. Rev. **77**, 388 (1950).
[17] H. Brooks, Phys. Rev. **83**, 879 (1951).
[18] P. P. Debye and E. M. Conwell, Phys. Rev. **93**, 693 (1954).

[19] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950), p. 264.

scattering probability used here,

$$1/\mu = A\langle Z^2\rangle_{Av} N_I \ln\left[1+\frac{B}{\langle Z^2\rangle_{Av} N_I^{\frac{1}{3}}}\right], \qquad (12)$$

where $A=\pi^{\frac{1}{2}}m^{*\frac{1}{2}}e^3 2^{7/2}\kappa^2(kT)^{\frac{1}{2}}$, and $B=(3kT/e^2)^2$. N_I is the total concentration of all ionized scatterers regardless of magnitude of charge, and $\langle Z^2\rangle_{Av}$ is the mean square of the charges,

$$\langle Z^2\rangle_{Av}=(\sum g_i Z_i^2 N_i)/N_I, \qquad (13)$$

the subscript i referring to those ions of the same charge magnitude.

It was shown in the previous section that, presumably because of different annealing kinetics of the interstitials and vacancies, the energy level model begins to break down appreciable exposure. Consequently it is useless to attempt to evaluate $\langle Z^2\rangle_{Av}$ on the basis of the model. However, one can determine approximately the product $\langle Z^2\rangle_{Av}N_I=N_I'$ from experiment by setting $\langle Z^2\rangle_{Av}N_I^{\frac{1}{3}}=N_I'^{\frac{1}{3}}$ in the logarithmic term. This approximation would tend to make N_I' too small,[20] but since for a maximum charge of two units $\langle Z_i^2\rangle_{Av}<4$, the error is probably less than a factor of 1.5. The ratio of N_I', the effective concentration of singly charged scatterers required to produce the observed scattering probability, to p_{77} was obtained by matching the experimenta curve of Fig. 7. The dashed curve shows the results obtained using $m^*=0.3m_0$ and $N_I'/p_{77}=8.72$. A reasonable agreement with regard to magnitude was obtained, but the calculated curve shows a distinct curvature which is not apparent in the experimental one.

In summary, we can only say that the nature of the scattering mechanism of bombardment-introduced defects is uncertain. Although certain features of the temperature and concentration dependence of the scattering probability are not consistent with charge-center scattering, the magnitude of the scattering probability is approximately that which would be expected if ~9 singly ionized defects were introduced for each ionized acceptor at 77°K. The reason for the large disparity in behavior of electron mobility to that observed for holes in bombarded Ge is not clear. More extensive experimental information is required to answer this question.

VI. LOWT-EMPERATURE BOMBARDMENTS

As in the case of n-type Ge,[1] the conductivity of a number of p-type specimens has been measured during exposure in the low-temperature facility of the graphite reactor at temperatures in the range −165° to −90°C. The variation of conductivity was also followed during various thermal cycles subsequent to exposure. The specimens investigated included those which were both originally p-type and those which had been converted to p-type by previous bombardment.

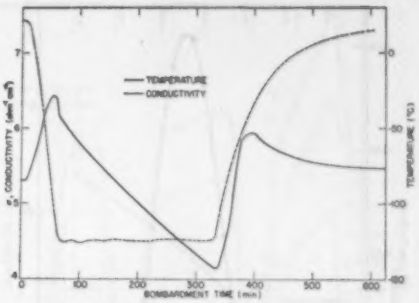


FIG. 8. Conductivity *vs* bombardment time for a p-type specimen during thermal cycling. The dashed curve is the temperature during bombardment.

For all specimens investigated, σ decreases during bombardment at these temperatures. In this respect their behavior is consistent with the Frenkel-defect energy-level model[2] and the analysis of the critical rate of hole removal at higher temperatures. For the value of ζ^* obtained above, it can be shown from Eq. (5) that $p^*\leq 8\times 10^{14}$ cm^{-3} for this temperature range. Since in all specimens $p>p^*$ at their respective temperatures of exposure T_e, a decrease in σ with bombardment is expected. Moreover, though not entirely conclusive, the dependence of the initial rate of decrease of σ on T_e and the initial value of σ or p_0 also appears to be in qualitative agreement with the model, the initial rate of decrease of σ being greater the lower T_e and the larger σ_0. In contrast to the behavior of σ during bombardment at constant temperature, which is apparently consistent with the chosen model, the variation of σ with temperature, during thermal cycling either during or subsequent to bombardment after a period of exposure at constant temperature, cannot be explained entirely on this basis. Similar results to those described below have also been observed in deuteron-bombarded Ge.[21]

The curve for σ *vs* bombardment time obtained for a typical run is shown in Fig. 8. The temperature throughout the run is shown by the dashed curve. Initially, as the sample is cooled, σ increases because of the temperature dependence of hole mobility. In the constant-temperature range (−135°C for this run), σ decreases monotonically with bombardment. At higher exposure temperature, the curvature in the constant-temperature region is much more pronounced. When the sample of Fig. 8 is warmed, σ increases rapidly to a

[20] There is another factor which has the opposite effect. The Conwell-Weisskopf scattering formula does not include screening effects. These are apt to become important even in this concentration range when $N_I>p$, thus decreasing the screening and increasing the scattering power of the ions (see reference 18). The error involved in the above approximation is expected to be compensated to some extent by the neglect of screening.

[21] Forster, Fan, and Lark-Horovitz, Phys. Rev. **91**, 229 (1953); See also J. H. Forster, thesis, Purdue University, 1953 (unpublished).

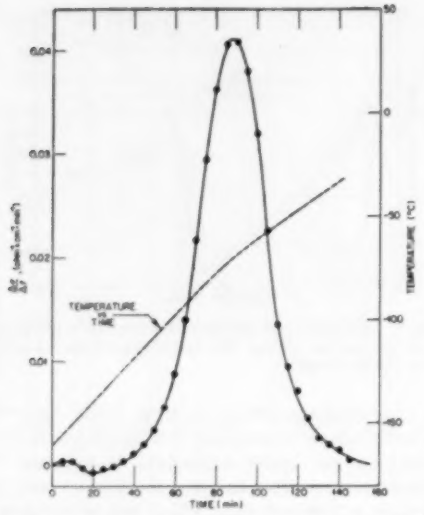


FIG. 9. The rate of conductivity change *vs* time after 640-min bombardment at −150°C. The dashed curve is the temperature.

maximum at −40°C and then decreases to a room-temperature value slightly higher than that at the start of the run. This increase in σ at room temperature is that expected to result from the same exposure period at room temperature. Similar behavior is observed if the warming is conducted after the reactor has been shut down.

Although the rapid increase in σ during warming may be due, all or in part, to thermal release of holes trapped in bombardment-produced defects, subsequent recooling of the specimen shows that all of the decrease in σ induced by low-temperature irradiation has annealed during the warming period. The rate of increase in σ during the heating cycle for a sample exposed for 640 min at −150°C is shown in Fig. 9; $\Delta\sigma/\Delta t$ is plotted as a function of time and the temperature during the cycle is indicated by the dashed curve. The curve shows a single maximum which is characteristic of a single-rate process. If the temperature dependence of mobility is suitably accounted for, the increase in σ during warming may be due to one or both of two processes: (1) the thermal release of holes from traps or (2) the annealing, by some mechanism, of the hole-trap defects which then release holes. Although recooling the specimen reveals that all of the decrease in σ, and hence the hole traps, has been removed during the heating cycle, the first process is by no means ruled out. A defect which behaves as a hole trap is stabilized by the approximate extent of the trapping energy when the trap is occupied. Consequently, it is possible that the annealing process takes place in two steps; the thermal release of holes from traps, followed by the annealing of those traps which have lost the stabilizing effect of the trapped hole. If such behavior is the case, the $\Delta\sigma/\Delta t$ *vs* time curve will not reflect the rate of annealing of hole-trap defects but, rather, the rate of thermal release of holes. This point will be considered in more detail in connection with pulse-annealing experiments which yield the necessary information for deciding between processes (1) and (2).

If specimens are irradiated for considerable periods at temperatures below −140°C, quite a different behavior of σ is observed during warming even though the form of the σ *vs* bombardment time curve is apparently the same as that obtained above this temperature. Initially, as the temperature begins to rise, there is a further sharp decrease in σ which persists until the temperature reaches ∼−120°C. Above this temperature σ rapidly increases in the manner shown in Fig. 8. This initial decrease in σ is evident in the initial portion of the curve of Fig. 9 in which $\Delta\sigma/\Delta t$ is negative.

In order to investigate the relaxation effects during warming in greater detail, several pulse-annealing experiments were carried out. The procedure used is as follows: A specimen was irradiated for the desired time at constant temperature. The reactor was then shut down and the specimen was warmed in place. The temperature was pulsed to a desired value, held constant for the desired time, and then returned to the reference value. The process is then repeated, pulsing to successively higher temperatures. This annealing technique has the advantage that the effect of a known thermal treatment on the specimen can be examined at a constant temperature, thereby eliminating complications associated with the temperature dependence of carrier concentration and mobility. The results of one such experiment, performed simultaneously on two p-type specimens after they had been irradiated for 18 hours at ∼−160°C, are shown in Fig. 10. The reference temperature during the run was −166°C. These curves indicated conclusively that σ is influenced by two distinct relaxation processes. After each pulse σ shows a definite decrease until −122°C is reached. After annealing at this temperature both specimens exhibit a σ which is lower than the initial value by more than an order of magnitude, the specimen with the lower initial σ showing the greater decrease. Pulsing to higher temperatures produces an irreversible increase which is characteristic of the annealing behavior of specimens irradiated above −140°C.

Inspection of Fig. 10 reveals that, besides the two irreversible processes, a reversible process, presumably thermal ionization of trapped holes, plays an important role in the σ *vs* t behavior during pulsing. The warm-up and cool-down behavior between pulses has been examined in detail. Typical $\log\sigma$ *vs* $1/T$ curves are shown in Fig. 11. After each pulse these curves are straight lines below −120°C and the slopes correspond to an apparent ionization energy of 0.10±0.01 ev for both specimens. This value of hole-trap depth is somewhat higher than that determined for the deep trap

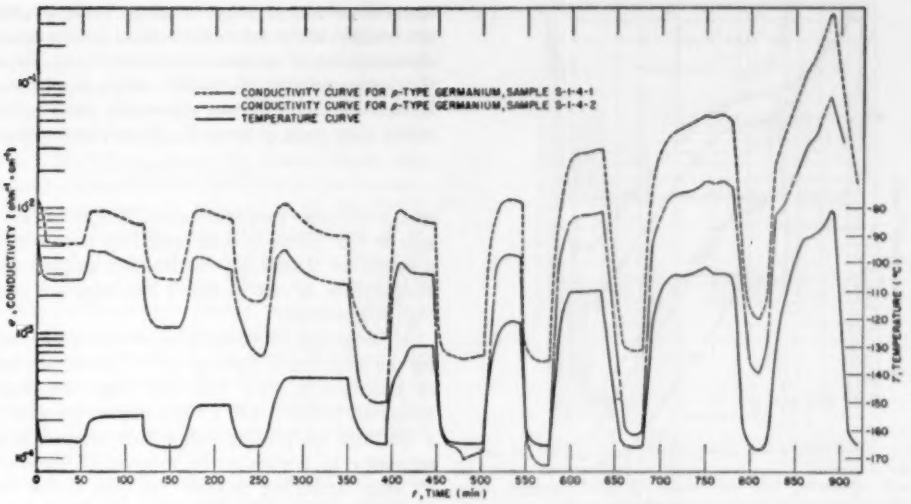

CONDUCTIVITY CURVE FOR p-TYPE GERMANIUM, SAMPLE S-1-4-1
CONDUCTIVITY CURVE FOR p-TYPE GERMANIUM, SAMPLE S-1-4-2
TEMPERATURE CURVE
σ, CONDUCTIVITY (mho-cm⁻¹)
T, TEMPERATURE (°C)
t, TIME (min)

FIG. 10. Conductivity vs time during a pulse-annealing run on two p-type specimens after 18-hour bombardment at -160°C. The temperature during the run is indicated by the solid curve.

from fitting the initial slope vs p_0 curve at dry ice temperature (Fig. 1, Sec. III). These results indicate that the defects responsible for the traps are able to stabilize themselves by $\sim$0.1 volt when they are occupied. In the temperature range under consideration this would correspond to appreciable stabilization. Returning to the question raised earlier in connection with the curve for $\Delta\sigma/\Delta t$ vs time (Fig. 9), in view of the possible stabilization effect and the evidence for the large reversible contribution to the pulse height, it would seem that the character of this curve is predominantly determined by the thermal release of trapped holes (process 1), at least in the low-temperature region covered by the pulse-annealing runs. Consequently, it is concluded that little can be learned concerning the annealing kinetics of hole-trap defects from curves of this sort.

We shall now examine the irreversible processes revealed by pulse annealing, considering first the higher temperature ($T>-120$°C) process. The irreversible increase in σ during warming is apparently quite similar to the process discussed in Sec. IV which occurs during prolonged room-temperature exposure or aging in the room-temperature range. The end result of both processes is a virtual removal of hole traps and a decrease in ionization energy of acceptors. The major difference in the two processes is the rate at which they occur. In the case of the low-temperature exposure, the relaxation is presumably inhibited by the low-thermal energy of the lattice and perhaps by the stabilization of hole traps as indicated above. Consequently, it is reasonable to expect that an appreciable amount of annealable disorder is stored during exposure at these temperatures. On warming, such a situation would produce rapid annealing over the temperature range in which the disorder becomes unstable. The nature of the relaxation process is not clearly understood. It is possible that the annealing or rearrangement of the disorder to a more stable configuration involves the removal of interstitials by the same mechanisms suggested above (Sec. IV). However, it is also possible that, although the effects of the process are similar to those at room temperature, the annealing mechanism is quite different. This is to be expected if, as has been suggested,[22] the disorder resulting from the collisions of fast neutrons with lattice atoms is highly localized in small regions and cannot be described by randomly distributed Frenkel defects. Such a configuration of disorder would be thermally unstable and would be expected to rearrange to one of greater stability as the thermal energy of the lattice increases. A careful study of the annealing kinetics of low temperature exposures is necessary to answer the questions posed here.

The effect of the irreversible process occurring below -120°C is essentially the opposite of the one just discussed in that a decrease rather than an increase in σ results. To explain this behavior it is necessary to postulate both the presence of minority-carrier traps and their removal by annealing at low temperature. There is convincing evidence that electron bombardment of p-type Ge at 78°K introduces electron traps.[23] Moreover, photoconductivity measurements as a function of temperature on p-type Ge bombarded at room

[22] Crawford, Cleland, Holmes, and Pigg, Phys. Rev. **91**, 243 (1953).

[23] Brown, Fletcher, and Wright, Phys. Rev. **96**, 834 (1954); Stoeckmann, Klontz, Fan, and Lark-Horovitz, Phys. Rev. **98**, 1535(A) (1955). Klontz, Pepper and Lark-Horovitz, Phys. Rev. **98**, 1535(A) (1955).

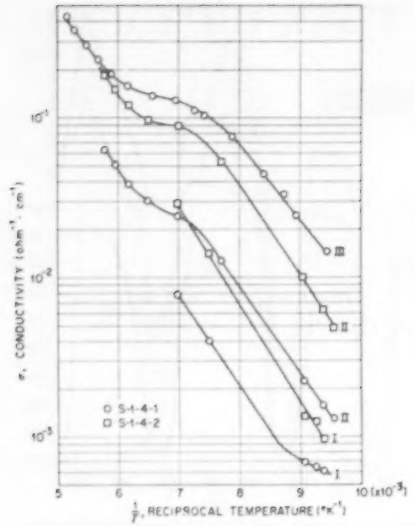


FIG. 11. Log conductivity *vs* reciprocal temperature. These curves are typical of those obtained after each pulse. Curves I, Curves II, and Curve III were obtained after the fourth, seventh, and eighth pulses respectively.

temperature with fast neutrons also indicates minority-carrier trapping.[24] Consequently, similar effects are expected in p-type Ge bombarded with fast neutrons at low temperature. Hence it is surprising that the curves of Fig. 11 indicate almost no temperature independent or nearly temperature independent component of σ at low temperature which would normally be expected to result from minority-carrier trapping under constant excitation (by γ radiation in the reactor). There is, however, a pronounced inflection in the $\log\sigma$ *vs* $1/T$ curves (Curves II and III) in the region of -120°C. Moreover, it is noted that the slopes of these curves are essentially the same both above and below the inflection which serves to off-set the curves toward lower values of σ at high temperatures. Minority-carrier trapping would indeed be expected to produce such behavior, provided that the concentration of hole traps were large compared to the total concentration of available holes.

This point can be illustrated as follows: Since an occupied hole trap is statistically equivalent to an acceptor, let us consider an acceptor level, equivalent to the level of hole traps, containing N states of which only N_a^0 are occupied by holes at 0°K ($N \gg N_a^0$). In the absence of exciting radiation and at sufficiently low temperatures ($p \ll N_a^0$), the law of mass action gives approximately

$$p \simeq N_a^0 K_a / N, \tag{14}$$

where the equilibrium constant K_a has the same form as Eq. (3). It is evident from Eq. (14), that in this approximation p is proportional to N_a: If Δn_t electrons are trapped under constant exciting radiation, an equal concentration of excess holes is produced. Because of the large number of vacant states in the acceptor level, these holes do not necessarily remain free, but rather they go to increase N_a. Hence, approximately,

$$N_a \simeq N_a^0 + \Delta n_t, \tag{15}$$

and p increases proportionately. The inflection in the $\log\sigma$ *vs* $1/T$ curves is then caused by a transition from a condition of minority-carrier trap saturation at low temperature to one in which the traps are empty at high temperature.

Although the above explanation adequately accounts for the inflection in the $\log\sigma$ *vs* $1/T$ curves, the intensity of radiation is such that the traps are essentially saturated below -120°C. In order to account for the σ decrease on pulsing below this temperature, it is necessary to postulate the removal of electron traps by some annealing process. Analysis of the behavior of σ during the third pulse (Fig. 10) indicates a first order relaxation process. The nature of the defects responsible for these electron traps is uncertain but, whatever the origin, some electron traps remain in the crystal at temperatures well above those at which the hole traps anneal.[24]

To summarize, low-temperature bombardment of p-type Ge decreases the conductivity, presumably by removal of holes by defects in a manner apparently consistent with the Frenkel-defect energy-level model. On warming after irradiation below -140°C, two distinct relaxation processes are observed: (1) a further decrease in σ which is virtually complete at -120°C and which may be associated with minority-carrier trapping phenomena and (2) the rapid annealing of the decrease in σ produced by bombardment at low temperature. The second process is similar in effect to annealing which has been observed in the room-temperature range. Extensive annealing-kinetics investigations in this temperature range are necessary before these processes can be understood. Such studies are now underway.

VII. DISCUSSION OF RESULTS

The distribution of defect-connected energy levels in bombarded p-type Ge, as indicated by an analysis of the initial behavior of specimens bombarded at room temperature and 195°K, is consistent with the Frenkel-defect model of James and Lark-Horovitz.[2]

The results for both n-type[1] and p-type material are summarized schematically in Fig. 12. The density of states in arbitrary units is plotted against energy over the forbidden energy band. Although the analysis yields discrete values for the defect levels, they are represented as bands in the diagram primarily because perturbations produced by neighboring defects are expected to broaden a set of discrete levels into a band.

[24] Cleland, Crawford, and Pigg (unpublished).

The bands of levels shown in Fig. 12(a) are as follows: Band A which is centered $\sim$0.2 ev below the conduction band, represents the shallow electron trap associated with the interstitial atom.[1] Band B is the second vacant level which, lying below the middle of the forbidden band, may either trap electrons or accept holes. This band, which lies 0.18 ev above the valence band, may be associated with either the interstitial or the vacancy, although the choice of γ's required to fit the 195°K data (Sec. III) indicates that the interstitial is responsible. Bands C and D are the low-lying occupied states which act as hole traps. These are presumably associated with the vacancy. Band C is placed $\sim$0.07 ev above the valence band. Although there is no direct information concerning band D, its depth as a hole trap is known to be considerably less than that of band C. Hence it is arbitrarily placed at 0.02 ev above the valence band. The last two bands are shaded to indicate that they are normally occupied.

The situation just described is expected for equal concentrations of vacancies and interstitials. If, as is indicated by room temperature annealing, an appreciable concentration of interstitials and the electrons originally associated with them are removed by some process, the distribution is expected to change to that indicated in Fig. 12(b). The result of such a process is to transform band C and a part of band D into acceptor states. This transformation would explain the low acceptor ionization energy values of converted n-type material after prolonged exposure and room-temperature aging (Fig. 5). Moreover, the ionization energy of 0.08 ev shortly after conversion (Curve I, Fig. 5) and the level indicated by the curves of Fig. 4 would be identified with band C in a manner consistent with this hypothesis.

It is possible, though the evidence is much more tenuous, to interpret the results of low temperature (<-90°C) exposure in the same manner. The disappearance of the bombardment induced decrease in σ on warming would then be explained by a transition from the situation depicted in Fig. 12(a) to that in Fig. 12(b). In view of alternative possibilities, however, much additional support for this view is required before it can be taken as conclusive.

Because of the consistency of results for both n- and p-type material with the model of James and Lark-Horovitz, it is indeed tempting to conclude that lattice defects introduced by fast-neutron bombardment can be described in terms of randomly distributed Frenkel defects. However, as pointed out previously,[1] because of the localized fluctuation in defect density expected to result from bombardment, this view should not be taken as conclusively established. Careful annealing studies of disordered specimens should help clear up this question. Comparison of the results of fast-neutron bombardment and electron bombardment would also be helpful since electrons with energies near the threshold value for displacing atoms from their normal lattice sites would be more likely to introduce randomly distributed defects.[25,26]

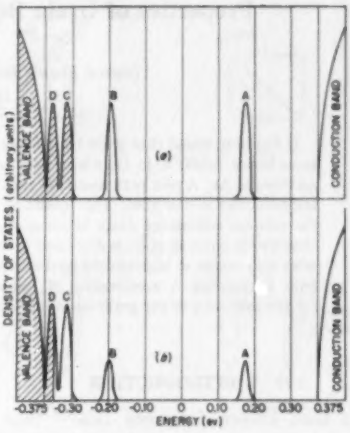


FIG. 12. A diagram of the distribution of energy levels produced by bombardment. Part (a) indicates the distribution expected for equal concentrations of interstitials and vacancies. Part (b) shows the effect expected to result from partial interstitial removal.

Another interesting point raised by these studies is the difference in the effect of bombardment on the mobility of holes as compared to that of electrons. Such a difference may well be associated with a nonrandom distribution of defects in that the charge density, and hence the electrostatic potential, associated with a disordered region may be a sensitive function of Fermi level. Such effects associated with dislocations in semiconductors have been recently considered by Read.[27] Consequently, a detailed study of the scattering process in Ge as well as annealing studies appears to be a valuable approach to a better understanding of the spatial distribution and detailed nature of bombardment-produced defects.

VIII. ACKNOWLEDGMENTS

We wish to thank Miss Louise Roth of Purdue University for furnishing the single crystals of Ge used in these experiments. In addition we are indebted to D. K. Holmes and H. C. Schweinler of this laboratory for many helpful discussions. The assistance of E. S. Schwartz in some of the measurements is gratefully acknowledged.

[25] E. E. Klontz and K. Lark-Horovitz, Phys. Rev. **86**, 643 (1952); see also E. E. Klontz, thesis, Purdue University, 1952 (unpublished).

[26] Brown, Fletcher, and Wright, Phys. Rev. **92**, 591 (1953).

[27] W. T. Read, Jr., Phil. Mag. **45**, 775 (1954); **45**, 1119 (1954); **46**, 111 (1955).

Properties of Grain Boundaries in Gold-Doped Germanium

A. G. TWEET
General Electric Research Laboratory, Schenectady, New York
(Received April 12, 1955)

It has been found that grain boundaries produce a path of relatively low electrical resistance at temperatures below ~100°K in Ge whose bulk resistivity has been made greater than 10^2 to 10^5 ohm-cm by the addition of Au. A new permanent-magnet apparatus for measuring the Hall effect in semiconductors at low temperatures is described. The results of measurement of the Hall effect as a function of temperature in Ge samples containing grain boundaries of known orientation are given. A peak in the Hall coefficient observed in some samples is discussed in terms of a crude model. Preferential conduction along grain boundaries also occurs in high-resistivity Fe-doped Ge, but has not been observed in crystals containing coherent twin boundaries. A modification of Bardeen's theory of surface states is used in an attempt to account for the existence of the grain-boundary conduction phenomenon.

I. INTRODUCTION

IT has been known for some time[1,2] that a grain boundary in *n*-type germanium introduces a layer of high-resistance material localized at the boundary, with a reverse saturation characteristic symmetrical with respect to the direction of a voltage applied across the grain boundary. Photovoltages[2] were also observed which reversed direction when the light spot crossed the boundary. When the bulk Ge was converted to *p*-type by heat treatment, the localized resistance at the grain boundary disappeared.[2] These facts have been interpreted by Pearson[2] to indicate the presence of an abnormally high concentration of acceptors at the grain boundary, producing a *p*-type layer.

A series of experiments were performed by the Purdue group[3] in which the breakdown voltage, dc conductance, and frequency-dependence of the impedance were all measured as a function of temperature on *n*-type Ge samples containing grain boundaries. These measurements were made with the voltage applied across the grain boundary. The results were interpreted in terms of a potential barrier model in which acceptor levels with energies in the forbidden energy gap are concentrated at the grain boundary, distorting the position of the band edges in energy much as in Bardeen's theory of surface states.[4]

Recently it has been found[5] that there is a path of relatively low electrical resistance along a grain boundary through a Ge crystal whose bulk resistivity has been made greater than 10^2–10^3 ohm-cm by the addition of Au.[6] The investigations described here were made possible by the characteristic high resistivity at low temperatures in bulk Au-doped Ge, which permitted the relatively slight conduction of the sheath of space-charge surrounding the grain boundary to be revealed.

Most of the experiments to be reported here were Hall effect and resistivity measurements on bicrystals of Au-doped Ge. A separate section (Sec. IV) includes short descriptions of other experiments.

II. EXPERIMENTAL

A. Apparatus

A small cryostat containing a permanent magnet was constructed for the low temperature Hall effect experiments. The outside brass vacuum can, *A* (Fig. 1), $3\frac{1}{2}$ in. o.d., 5 in. high, was made in two parts, with a flange and gasket seal, *B*. A stainless steel tube, *C*, served both as a vacuum lead and also to suspend the cryostat from the Al lid, *D*, of a straight-sided commercial bench Dewar vessel, (not shown), which contained either liquid N_2 or liquid H_2. The bottom of the

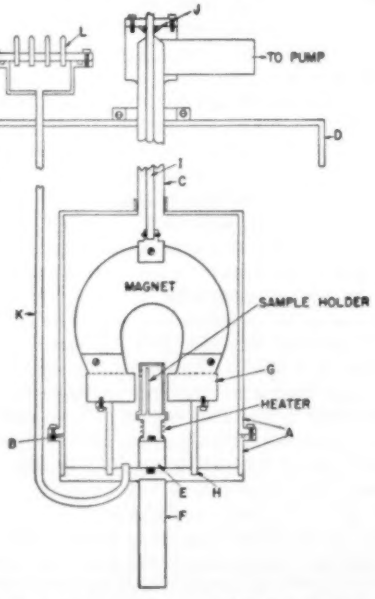


FIG. 1. The permanent-magnet vacuum Hall effect apparatus. The Dewar vessel surrounding the cryostat is not shown.

[1] K. Lark-Horovitz, National Research Council Report NDRC-14-585, 1945 (unpublished).
[2] G. L. Pearson, Phys. Rev. **76**, 459 (1949).
[3] Taylor, Odell, and Fan, Phys. Rev. **88**, 867 (1952).
[4] J. Bardeen, Phys. Rev. **71**, 717 (1947).
[5] A. G. Tweet, Phys. Rev. **96**, 828 (1954).
[6] W. C. Dunlap, Jr., Phys. Rev. **91**, 1282 (1953).

Cu sample holder had a heater of No. 30 Constantan wire, resistance 50 ohms. A Cu radiation shield in fairly good thermal contact with the heater, enclosed the sample. The sample holder was electrically insulated from the sample by a thin sheet of mica glued on with *GE*7031 cement. The stainless steel heat leak, *E*, connected the sample holder to a Cu post, *F*, which depended into the cooling bath.

The Alnico *V* magnet, with its soft iron pole pieces, *G*, gave a flux of ~1400 gauss at the center of a gap $\frac{5}{8}$ in. across and $\frac{1}{8}$ in. square. The magnet could be rotated in a race, *H*, by means of the rod, *I*, which passed through an O-ring seal, *J*, at room temperature. The leads for the sample, heater and thermocouple came up through the tube, *K*, to a header, *L*. The gasket for the low-temperature vacuum seal was a torus of 0.050-in. In wire. A Welch Duoseal forepump was used to evacuate the cryostat, since only a rough vacuum was required to isolate the sample sufficiently from the cooling bath. Heat leaks from above were reduced by gluing a 1.5-in. thick plug of Styrafoam to to the under side of the lid, *D*. Openings in the lid, not shown in the drawing, provided a means of filling the Dewar flask. The lead wires were all 0.010-in. Cu Formex coated. The thermocouple wires were continuous to avoid stray thermal and contact emfs. They were brought out of the header through a seal of DeKhotinsky cement. The measurements of voltage were made with a Minneapolis-Honeywell vibrating reed electrometer. Sample temperature was measured by a (AuCo)−(AgAu) thermocouple,[7] read on a Rubincon Portable Precision Potentiometer. A heater power of ~5 watts raised the temperature of the sample to 200°K when the cooling bath was liquid N_2. When the apparatus has been precooled with liquid N_2, 2 liters of liquid H_2 are sufficient to cool it and maintain the sample at ~27°K for at least 6 hours.

The functioning of the apparatus, which has been in operation for over 8 months, has been checked against samples of Ge measured by W. C. Dunlap, Jr., and the results agree within 25 percent.

B. Procedure

Bicrystals of *n*- and *p*-type bulk Au-doped Ge were pulled from the melt using two [100] seeds bound symmetrically on either side of a Mo wedge. If the axis of rotation of the angle β between the two seed crystals is a [110] direction, then the resulting ingot is usually as shown in Fig. 2(A). β was approximately 10° in all the work to be reported here except when stated otherwise. If a wafer is cut from the ingot, as shown by the dotted lines in Fig. 2(A), the path of the grain boundary across the wafer seen from above is almost always as shown in Fig. 2(B) with the angle γ between 20° and 60°. However, if the axis of rotation of the angle β is a [100] direction, then the grain boundary very seldom departs from planarity, and the appearance of a wafer cut from such an ingot is usually as shown in Fig. 2(C).

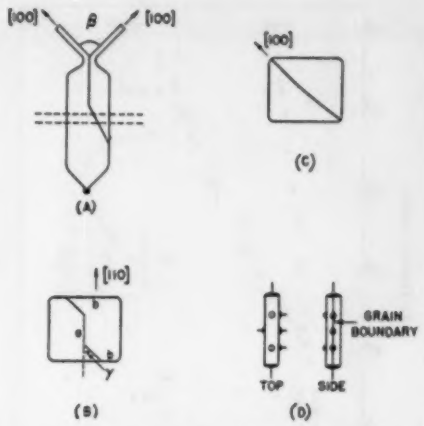


FIG. 2. (A) Schematic drawing of a bicrystalline ingot. (B) A wafer cut from the ingot in (A) along the dotted lines, and viewed along the ingot axis. (C) Same as (B) except that the axis of β is the [100] direction instead of [110]. Note that the grain boundary in (C) does not have the same orientation as section *b* of the grain boundary in (B). (D) Schematic drawing of Hall wafer containing a grain boundary.

For the Hall effect measurements, wafers ~0.10 in. ×0.15 in. ×0.40 in. were cut from the Ge bicrystals so as to leave the grain boundary exposed, and In contacts fused on, as in Fig. 2(D), under H_2. All samples were etched carefully after putting on the contacts. The resistance was measured at liquid N_2 temperature between the contacts placed on the top surface of the wafer (open circles) and each of the contacts on the grain boundary to see that the latter were not shorting to one another through a conducting surface layer on the high-resistance Au-doped Ge. (Such shorting effects are often observed in high-impedance samples but can be avoided by careful handling after etching.) The samples were then attached to the sample holder with *GE*7031 cement which, when dry, provides moderately good thermal contact and excellent electrical insulation. The thermocouple was soldered directly to one of the contacts initially used to check for surface leakage and the apparatus closed, evacuated, and cooled.

Voltages were checked for linearity as a function of current magnitude and direction of flow through the whole temperature range of the measurements.

III. RESULTS AND DISCUSSION

A. Preliminary

In Fig. 3 is shown a schematic drawing of a wafer cut from a bicrystal of *p*-type Au-doped Ge whose orientation is that of Fig. 2(B). Solid circles indicate the position of fused In contacts on one surface of the

[7] The author is indebted to J. F. Youngblood of this laboratory for the thermocouple.

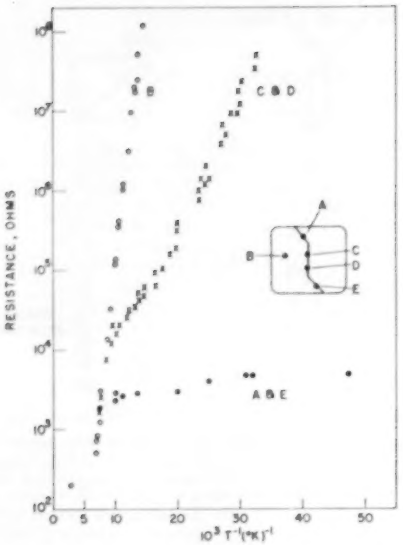


FIG. 3. Preliminary measurement showing qualitatively the grain boundary conduction phenomenon in Au-doped Ge. The inset gives the disposition of contacts on a sample of the orientation shown in Fig. 2(B). Curve B is of the form $e^{+\varphi/kT}$, where $\varphi=0.15$ ev, characteristic of the lower Au acceptor state in Ge. Curves labeled C and D, and A and E show the short-circuiting effects of the grain boundary.

wafer. The other surface has had an overall contact of In fused to it. The data points represent resistance measurements made with a Volt Ohmyst ohm-meter between the large area contact and the dot contacts in turn, through the bulk Ge (contact B), and through the various sections of the grain boundary (contacts A, C, D, and E). Curve B is an exponential of the form $e^{+\varphi/kT}$, where $\varphi=0.15$ ev, the energy associated with the lower gold acceptor state in Ge.[6] The curve labeled C and D has a slope corresponding to a φ of 0.04 ev. The extreme short circuiting effect of the conducting sheath concentrated at the grain boundary seen on the sections of crystal labeled A and E occurs for a considerable range of angle γ, defined in Fig. 2(B). On this wafer, $\gamma\sim40°$. These data, taken on a single wafer, together with similar data on many other samples, constitute evidence that the electrical properties of the grain boundary are a function of the orientation of the boundary, since all other conditions in the growth of the ingot are presumably identical.

The data also indicate the necessity for avoiding the inclusion of sections of grain boundary containing jogs when preparing a Hall wafer. All the Hall effect measurements were made on wafers in which the grain boundaries were believed to be free of jogs, as was inferred by viewing the straightness of their peripheries under a low-power (100×) microscope. However, the presence of internal irregularities in the wafers cannot be ruled out.

It is not uncommon for very-small angle grain boundaries to spread out from the main boundary as tributaries from a river, sometimes in considerable profusion. These boundaries all separate crystals of misorientation less than 1° as measured by x-ray Laue photographs,[8] and can be revealed only by rather careful etching. About five wafers containing these tributary boundaries were measured by the technique used in obtaining the data for Fig. 3. No short-circuiting of the bulk Au-doped Ge because of the presence of the very small angle boundaries was found in these wafers up to resistances of 100 megohms. However, no wafers containing such tributary boundaries were used for Hall effect measurements.

B. Resistivity and Hall Coefficient

In this section data are presented on the reduced resistivity and reduced Hall coefficient of two wafers of p-type Au-doped Ge containing grain boundaries. The sign of the carriers in the grain boundary con-

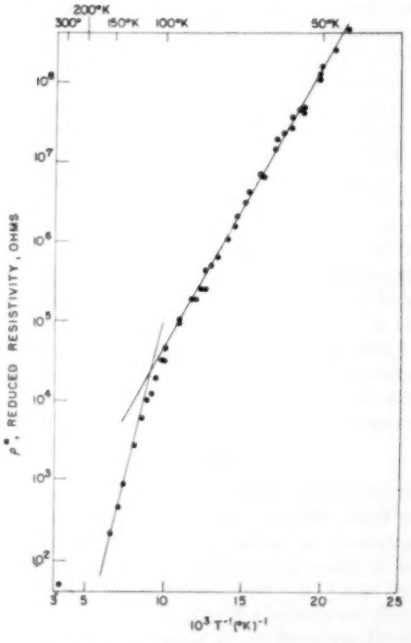


FIG. 4. Reduced resistivity, ρ^*, for a sample of Au-doped Ge containing a grain boundary such as a in Fig. 2(B). Above 100°K, ρ^* is of the form $e^{+\varphi/kT}$, where $\varphi=0.15$ ev, characteristic of the lower Au acceptor state in Ge. Below 100°K, it is of the same form, with $\varphi=0.07$ ev, representative of φ's observed for other samples with the same orientation of grain boundary. Above ~125°K, the true resistivity of the bulk Ge may be obtained by multiplying ρ^* by the sample thickness, $t=0.25$ cm.

[8] The author wishes to thank Mrs. A. Cooper of the Metallurgy Research Department for these x-ray measurements.

duction was p-type for both wafers. This has been the case for all samples studied, even when the bulk Ge was n-type. The significance of the use of the term "reduced Hall coefficient" is the following. For a normal Hall wafer,[9]

$$R = 10^8 V_H t / HI,$$

where the Hall coefficient, R, is in cm³/coulomb when the Hall voltage V_H is in volts, the sample thickness t, is in cm, the magnetic field, H, is in gauss and the current, I, is in amperes. This formula is applicable to a homogeneous sample when the sample thickness is a measurable quantity. However, in the case of the grain boundary space-charge layer conduction in Au-doped Ge, the thickness of the conducting layer can only be estimated and the density of carriers is almost certainly a function of the perpendicular distance from the grain boundary. Consequently, it is difficult to see the physical significance of R. But if one defines R^*, the reduced Hall coefficient, as $R^* \equiv 10^8 V_H / HI$, then all the quantities on the right side of the equation are measurable. R^* is the reciprocal of the total charge/cm² of sample, provided the volume density of charge is a function only of the perpendicular distance from the grain boundary, and the mobility is not a function of position in the Hall wafer. In p-type Au-doped Ge, at room temperature R^*t is the Hall coefficient of bulk Ge, since now the fraction of the current carried by the grain boundary space-charge sheath is not detectable. At sufficiently low temperatures, R^* is the reciprocal of the total charge/cm² of grain boundary associated only with the grain boundary conduction. At intermediate temperatures, when comparable fractions of the current are carried by bulk Ge and grain boundary, a quantitative interpretation of R^* is difficult. However, in this temperature range there are other more fundamental difficulties in interpreting the measurement, as will become evident, and hence no detailed treatment at this time seems profitable.

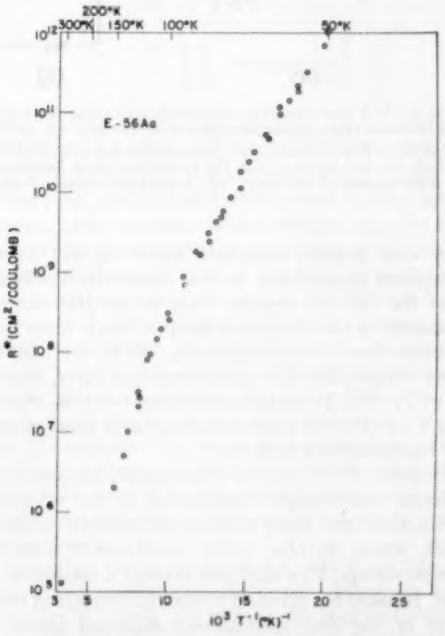


FIG. 5. Reduced Hall coefficient, R^*, for the sample of Fig. 4. The true Hall coefficient of the bulk Ge above ~125°K may be obtained by multiplying R^* by t=0.25 cm.

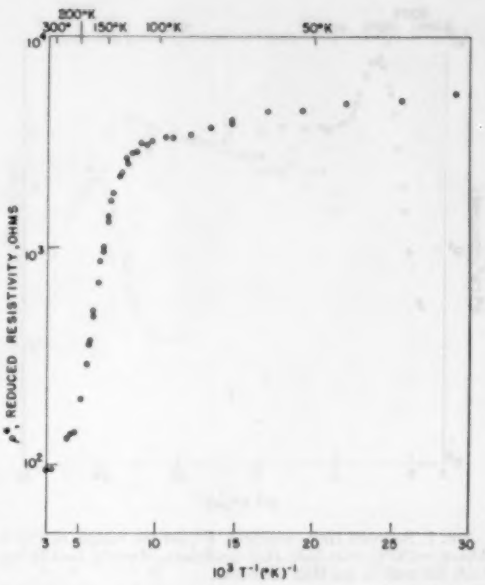


FIG. 6. Reduced resistivity for a sample of Au-doped Ge containing a grain boundary as in Fig. 2(C). Above ~150°K, true bulk resistivity $\rho = \rho^* t$, t=0.21 cm.

The problem of the unknown thickness of space-charge sheath at the grain boundary and its nonuniform density also occurs in the calculation of sample resistivity. The reduced resistivity, $\rho^* \equiv V_s w / I$. It is in ohms when V_s, sample voltage, is in volts and the sample current, I, is in amperes. The quantities w and l are the sample width and separation of the voltage probes, respectively. At temperatures where the grain boundary charge sheath carries a negligible fraction of the sample current, $\rho^* l$ is the true sample resistivity.

The results of measurements of the reduced resistivity, ρ^*, and reduced Hall coefficient, R^*, on a sample cut from a bicrystal of Au-doped Ge containing a section of grain boundary such as a in Fig. 2(B) are shown in Figs. 4 and 5. The steeper slope at temperatures above 100°K corresponds to a φ of 0.15 ev and is characteristic of the lower Au acceptor level in Au-

[9] W. Shockley, *Holes and Electrons in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950), pp. 211–215.

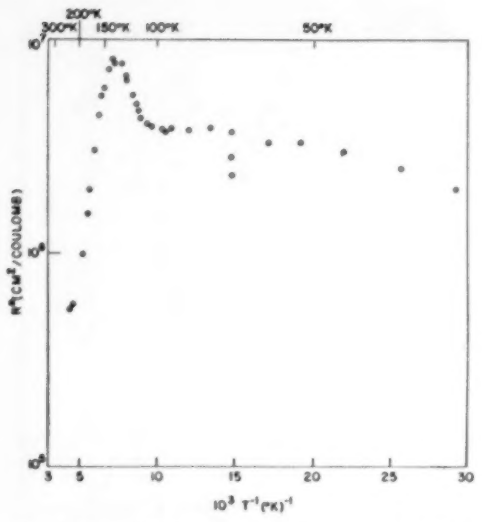


FIG. 7. Reduced Hall coefficient, R^*, for the sample of Fig. 6. Above ~150°C, true bulk Hall coefficient, $R=R^*t$, $t=0.21$ cm. Note the peak in the Hall coefficient.

doped Ge. The curve changes slope to 0.07 ev below about 80°K, which is representative of values of φ ranging from 0.03 ev to 0.07 ev seen in similar samples taken from other ingots.

Plots of reduced resistivity and reduced Hall coefficient for a sample cut from an ingot whose grain boundary is as shown in Fig. 2(C) are given in Figs. 6 and 7. Note that even though the resistance of this sample is similar in temperature dependence to that of the sections marked A and E on Fig. 3, these grain boundaries do not have the same orientation.

The peak in the reduced Hall coefficient in Fig. 7 has been observed in all five samples taken from ingots similar to the one whose measurements are reproduced here. It is believed to be characteristic of all grain boundaries with this orientation.

C. Discussion of Hall Effect Measurements

The proper interpretation of the Hall effect data is uncertain, quite aside from the difficulty connected with nonuniform carrier density mentioned in Sec. III. B. This uncertainty is made particularly evident by the marked maximum in the Hall coefficient in Fig. 7. It is extremely unlikely that the carrier concentration, no matter how it is distributed across the thickness of the sample, should increase as the temperature decreases, which is the conclusion one is led to if he assumes that $R=1/nec$, the usual interpretation.[9] Consequently further interpretation of the data is necessary.

Consider an idealized model of a Hall wafer of Au-doped Ge containing a grain boundary as composed of two homogeneous p-type regions, such as in Fig. 8(A). Let the first region represent the bulk Ge in which the resistivity increases rapidly with decreasing temperature and the mobility is limited by thermal vibrations, increasing as the temperature decreases. In the second region, which we will identify with the very much thinner grain boundary space-charge sheath, assume that the resistivity remains relatively constant with decreasing temperature and that the mobility is limited by impurity and dislocation scattering, decreasing as the temperature decreases.[10]

When a current, I, flows in such a composite sample in the presence of a magnetic field, H, [Fig 8(A)], in general a circulating current will flow through the Hall contacts between regions 1 and 2 in the plane of the figure, because of the difference between the Hall mobilities, and hence the Hall voltages, characteristic of the two regions.

Figure 8(B) shows a crude equivalent circuit for this situation. V_1 and V_2 are the Hall voltages which would be developed across regions 1 and 2, respectively, if

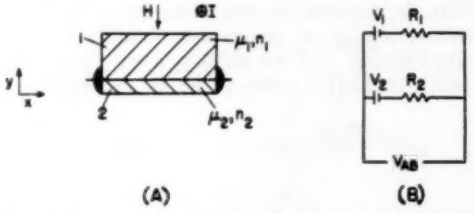


FIG. 8. (A) A crude model of a Ge sample containing two regions with different temperature-dependences of mobility (μ_1 and μ_2) and carrier concentrations (n_1 and n_2). Region 1 is identified with the bulk Ge and region 2 with the conducting layer localized at the grain boundary (see text). (B) Equivalent circuit of model in (A).

they were isolated from each other. R_1 and R_2 are resistances proportional to the (resistivity/thickness) ratio for the two regions. V_{AB} is the Hall voltage measured in the composite sample. From room temperature down to approximately 200°K, the bulk Ge short circuits the thin grain boundary layer, $R_2 \gg R_1$ and $V_{AB}=V_1$. At temperatures below ~100°K, $R_1 \gg R_2$ and $V_{AB}=V_2$: the grain boundary layer short circuits the high-resistivity bulk Ge.

However, $V_2<V_1$ at low temperatures because of the different temperature dependences of the scattering mechanisms and hence at some intermediate temperatures, where $R_1 \approx R_2$, under conditions of constant sample voltage, V_{AB} must pass through a maximum.

H. Brooks[11] has given an elegant mathematical treatment of the ideas qualitatively described above, as follows. Referring to Fig. 8(A), remove all restrictions

[10] D. L. Dexter and F. Seitz, Phys. Rev. **86**, 964 (1952).

[11] H. Brooks (private communication). The author is much indebted to Professor Brooks for permission to publish this summary of his analysis.

on the y-dependence of μ and n, the Hall mobility and carrier concentrations, respectively. Let $J_x(y)$ be the current density of the circulating current described above, which may now be a function of position in the direction perpendicular to the grain boundary. From the continuity equation,

$$\int_0^t J_x(y)dy=0,$$

where t is the sample thickness, since there are no sources or sinks in the wafer.

But

$$J_x(y)=\sigma(y)[E_H(y)-V_{AB}/w],$$

where $\sigma(y)$ is the conductivity of the Hall wafer, a function of position in the y direction as well as temperature. $E_H(y)$ is the Hall field, V_{AB} is the measured Hall voltage, and w is the sample width. Substituting for the Hall field,

$$E_H(y)=\mu_H(y)HE_x/c,$$

where E_x is the drift field and $\mu_H(y)$ is the Hall mobility, one obtains

$$V_{AB}=\frac{IH}{ec}\frac{\int_0^t n(y)\mu_H(y)\mu_D(y)dy}{\left[\int_0^t n(y)\mu_D(y)dy\right]^2}, \qquad (1)$$

where μ_H and μ_D, the drift mobility, are not assumed to be the same. Equation (1) describes the circuit of Fig. 8(B) in the limit of two homogeneous regions.

The peak in Hall coefficient in sample E-46A2 occurs because under constant applied voltage, only the Hall voltage is temperature-dependent below ~100°K and its peak controls the peak in the Hall coefficient. On the other hand, there is *no* peak in the Hall coefficient of sample E-56Aa because the rapid decrease with falling temperature in sample current at constant sample voltage (see Fig. 4) completely overshadows the decrease in Hall voltage and so the Hall coefficient continues to rise as the temperature decreases. Equation (1) suggests that the continued decrease in Hall coefficient for sample E-46A2 as the temperature drops below 60°K may be evidence for the existence of layers of different mobility in the grain boundary space-charge sheath.

From Eq. (1) it can be shown that if the form of $n(y)$ in the space-charge layer surrounding the grain boundary is known, the ratio of the Hall mobilities in the grain boundary layer and the bulk Ge at the temperature of the maximum in the Hall coefficient can be calculated from the experimental data.

Figure 9 shows Hall mobilities for the two samples whose resistivities and Hall coefficients are presented

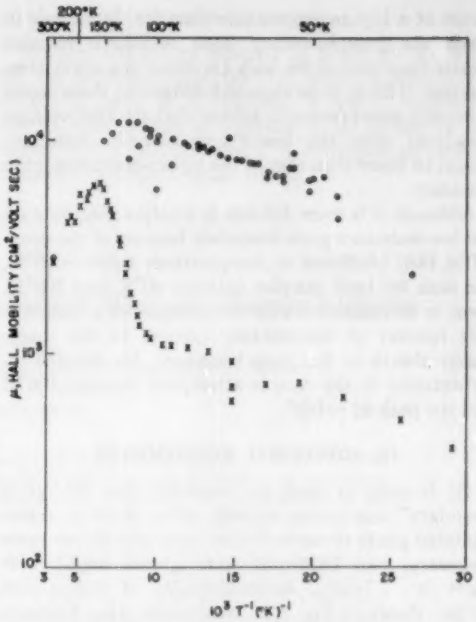


FIG. 9. Hall mobilities for samples of Figs. 4 and 6. ○ Sample of Fig. 4; × sample of Fig. 6. Note that for the calculation of the Hall mobility it is not necessary to know the thickness of the conducting layer.

in Figs. 4–7. The mobilities were calculated from the formula[9]

$$\mu=\frac{10^8}{H}\frac{V_H}{V_x}\frac{s}{w},$$

where μ is in cm²/volt sec if H is in gauss. V_H is the Hall voltage, V_x, the sample voltage, and s and w are, respectively, the separation of the resistivity and Hall probes. Both sets of data can be divided into three regions according to temperature. There is a high-temperature range in which the mobility is essentially that of the bulk Ge, which is carrying all the current. There is a low-temperature range in which all the current is confined to the space-charge layer at the grain boundary and the mobilities are characteristic of the carriers in this layer, subject to the limitations of interpretation imposed by the above discussion of the modification in Hall voltage encountered in composite samples. There is also an intermediate-temperature range, coincident with the neighborhood of the maximum in the Hall coefficient in one of the samples, in which the Hall voltage and mobilities have their peaks, reflecting the transition from one kind of conduction to the other. It is evident that in the sample with the low grain-boundary resistance, the shorting effect on the Hall voltage (or mobility) of the bulk Ge due to the presence of the grain-boundary space-charge layer

occurs at a higher temperature than for the sample in which the grain-boundary layer resistance remains greater than that of the bulk Ge down to a lower temperature. This is to be expected. However, there seems to be no *a priori* reason to believe that the Hall voltage associated with the low-resistance grain boundary should be lower than that of the higher-resistance grain boundary.

Although it is more difficult to interpret the data on the low-resistance grain boundary because of the drop-off in Hall coefficient at temperatures below ~60°K, the data for both samples between 60°K and 100°K, seem to be consistent with the presence of a considerable number of low-mobility carriers in the space-charge sheath at the grain boundary. No detailed interpretation of the data is attempted between 100°K and the peak at ~160°.

IV. ADDITIONAL EXPERIMENTS

(1) In order to check the possibility that the "grain boundary" conduction actually takes place in a precipitated phase of Au-rich alloy localized at the grain boundary, two bicrystals were grown doped with Au^{198} (β^-, 2.7-day). Autoradiographs of wafers such as that shown in Fig. 2(B) were made using Eastman autoradiographic plates "*t*-coating," but no additional fogging of the plates at the position of contact with the grain boundary was observed. The resolution of this method is very poor, however, and the experiments do not preclude the possibility that preferential segregation of the Au, as well as other impurities, occurs at the grain boundary.

(2) Another possibility is that preferential grain boundary diffusion[12] takes place when the fused In contacts are being put on, and that the rapidly diffused In is really responsible for the grain boundary conduction. However, by smearing Al-Ga paste on the ground surface of one face of a wafer and probing the opposite face with a point, preferential grain-boundary conduction could be detected quite readily on wafers which had had no heat treatment after growth.

(3) As mentioned previously, most of the experiments reported here have been performed upon bicrystals with angle β as defined on Fig 2(A) of 10°. However, crystals with twin boundaries have been grown and in these it was impossible to detect the presence of the boundary by electrical resistance measurements up to resistance of the order of 100 megohms.

(4) To examine the width of the conducting layer surrounding the grain boundary in a wafer of Au-doped Ge, an In contact was fused to the grain boundary region of a wafer cut from a bicrystal of Au-doped Ge. The resistance was then measured between this contact and a 0.001-in. W point set down at regularly spaced intervals on the surface with a micromanipulator. The measurements were made at liquid N_2 temperature with the surface of the wafer continuously immersed in liquid N_2. The spatial resolution of the experimental arrangement was improved by a factor of 10 by cutting the wafer at an angle of ~5° to the grain boundary and then probing across the etched surface of the wafer. The point was lifted from the Ge when moving it from place to place on the wafer to avoid scratching the surface. Considerable care was exercised to minimize effects associated with variations in point pressure. The width of the conducting layer is apparently not more than ~2μ, an experimental upper limit set by the width of the W point.

(5) It is of interest to know whether or not the phenomenon of grain boundary conduction is confined to crystals doped with Au. As a check, bicrystals of Fe-doped Ge[13] were grown, and it was found that the short-circuiting effects of a grain boundary occurred here as well.

(6) Hall measurements were made on pairs of wafers cut from the same ingot with their long directions at right angles to each other in a search for directional dependence of the mobility in the grain boundary conduction, but with no conclusive evidence that such a dependence exists.

V. CONCLUSIONS

The results of the measurements reported here are consistent with the existence of an array of acceptor levels localized at the grain boundary in a Ge bicrystal, with energies near the valence band. Electrons from the valence band may be trapped by these centers and the resulting holes confined to a space-charge layer surrounding the array of negatively charged acceptor levels. The difference in the temperature dependence of the reduced resistivity for the samples described here may possibly be related to a dependence of the energy of the acceptor levels with respect to the valence band upon the orientation of the grain boundary.

It is not clear what the source of the grain-boundary acceptor levels is. They may be related to the distorted and unsaturated valence forces discussed by Read,[14] in which case it is quite plausible that their energy should depend upon whether the dislocations in the grain boundary are pure edge or have a sizable screw component. On the other hand, impurities trapped at the grain boundary could be responsible for its acceptor action. This seems somewhat unlikely, however, in view of the available evidence concerning the binding energy of impurities to dislocations.[15,16]

[12] D. Turnbull and R. E. Hoffman, Acta Metallurgica 2, 419 (1954).

[13] W. W. Tyler and H. H. Woodbury, Phys. Rev. 96, 874 (1954).

[14] W. T. Read, Jr., Phil. Mag. 45, 775 (1954).

[15] A. H. Cottrell, *Relation of Properties to Microstructures* (American Society for Metals, Cleveland, 1952), p. 131.

[16] The author is indebted to W. T. Read and to D. Turnbull for stimulating conversations on this subject.

As might be expected, the temperature dependence of the mobility of the carriers in the grain-boundary space-charge sheath is such as to indicate that the principle scattering mechanism is impurity or defect scattering However, the marked difference between the mobilities (Fig. 9) of samples with different orientation of the grain boundary is not understood.

ACKNOWLEDGMENTS

It gives the author pleasure to thank C. W. Wolff for the construction of the Hall effect apparatus. He wishes to thank W. C. Dunlap, Jr. and L. Apker for their advice and encouragement in all phases of the work; and E. O. Kane and C. J. Gallagher for several valuable discussions of the results.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Hot Electron Problem in Semiconductors with Spheroidal Energy Surfaces

MOTOICHI SHIBUYA
Physics Department, University of Illinois, Urbana, Illinois, and Electrotechnical Laboratory, Nagata-cho, Chiyoda-ku, Tokyo, Japan

(Received April 22, 1955)

Shockley has derived the expression $E_c = (1.51)C_l/\mu$ for the critical electric field strength E_c at which the current-voltage relation in a semiconductor departs from Ohm's law. Here C_l is the velocity of longitudinal phonons and μ the conductivity mobility of the carrier. Experimental values of E_c for Ge and Si are from two to four times larger than those predicted by this formula. We therefore have extended the theory to take account ellipsoidal energy surfaces in the Brillouin Zone and scattering by shear modes of vibration. The effect of the more general effective mass tensor is to raise the theoretical value of E_c by a factor of about 2 for n-type Ge and 1.3 for n-type Si, whereas shear mode scattering lowers E_c by a factor that is between 1 and the value of the ratio of the velocity of transverse modes to that of longitudinal modes. Moreover, E_c should vary with the direction of the current. The present study still fails to close the gap between theory and experiment: the remaining discrepancy is possibly the result of neglecting intervalley scattering.

INTRODUCTION AND SUMMARY

THE deviation from Ohm's law in semiconductors such as Ge and Si in strong electric fields has been studied both theoretically[1–7] and experimentally[8,9] by many investigators. This problem has been called the "hot-electron" problem by Shockley because the average kinetic energy of the electrons in strong fields becomes larger than the usual thermal energy, so that the "electron temperature" becomes larger than the lattice temperature. There are, however, quantitative discrepancies between theory and experiment in some respects, such as in the value of the critical field E_c where the deviation from Ohm's law occurs. The experimental values in Ge and Si are about twice to four times larger than those obtained from the formula

$$E_c^s = 1.51 C_l/\mu. \tag{1}$$

The superscript s on E_c indicates that this value is obtained from the scalar effective mass theories, C_l is the longitudinal velocity of the phonons, and μ is the weak-field conductivity mobility of electrons.

Previous calculations have assumed spherical energy surfaces in the Brillouin zone. It is now believed that for n-type Ge and Si one must use spheroidal surfaces in discussing the electrical properties. Further, as has been pointed out by Adams[10] and by Herring,[11] shear waves are as important as pure compressional waves in the lattice scattering of electrons in these crystals. Therefore we have attempted to solve the hot-electron problem with the spheroidal model, including the effect of shear waves. A number of drastic approximations are made in solving the Bloch equation, and we finally get the expression for the critical field in the following form:

$$E_c = A S E_c^s, \tag{2}$$

where the factor A is an anisotropy factor and the other factor S results from the effect of the shear waves. When we neglect the effect of the latter, S equals 1, and when the mass component ratio $r = m_2/m_1$ is equal to 1, A becomes independent of the direction of the current vector and is also 1. The larger the departure of r from 1, the larger is the value of A. The factor S is a complicated function of the mass ratio r, the ratios of the characteristic coefficients of the deformation potentials of shear waves and of pure compressional waves, and the ratio $C = C_t/C_l$ of the transverse and longitudinal velocities of the acoustical modes of lattice vibration. However, the value of S is almost always between C and 1. This fact may be qualitatively understood, because both

[1] W. Shockley, Bell System Tech. J. **30**, 990 (1951).
[2] Druyvesteyn, Physica **10**, 61 (1930).
[3] F. B. Pidduck, Proc. Lond. Math. Soc. **15**, 89 (1915).
[4] B. Davidow, Physik. Z. Sowjetunion **8**, 59 (1936).
[5] L. Landow and A. Kompaniez, Physik. Z. Sowjetunion **6**, 163 (1934).
[6] E. Guth and J. Meyerhofer, Phys. Rev. **57**, 908 (1940).
[7] J. Yamashita and M. Watanabe, Repts. Inst. Sci. Tech., Tokyo Univ. **6**, 111 (1952) (in Japanese).
[8] E. J. Ryder and W. Shockley, Phys. Rev. **81**, 139 (1951).
[9] E. J. Ryder, Phys. Rev. **90**, 766 (1953).
[10] E. N. Adams II, "Notes on the energy band structure of Ge and Si" September, 1954 (unpublished).
[11] C. Herring, "Transport properties of a many-valley semiconductor" September, 1954 (unpublished).

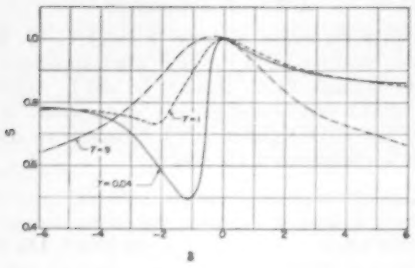


FIG. 1. S *versus* δ for $r=0.04$, 1 and 9.

transverse and longitudinal modes of vibration contribute to shear-wave scattering, so that the effective average velocity for lattice scattering is between C_t and C_l. We already found scalar conductivity for cubic crystals in weak electric fields without magnetic fields even in the spheroidal approximation,[12] because of the cubically symmetric distribution of spheroids. In the hot-electron problem, on the contrary, anisotropic effects do not vanish with the same spheroidal model. Thus the factor A depends not only upon the model but also upon the direction of the current vector. In addition, when the current is not parallel to the symmetry axes of the cubic crystal such as the (100), (110), and (111) axes, there appears a small induced transverse field. Although the angle ϵ between the current and field vectors is very small, about a few degrees, its maximum value is about one hundred times larger than the mosaic angle in the case of Ge. There have been no experimental observations of the transverse voltage and the directional dependence of the critical field arising from the anisotropy of the spheroids. However, we believe that this effect should be detectable.

We have neglected intervalley scattering and scattering due to impurities and to optical modes of vibration and made many approximations and simplifications in this calculation. Therefore it is not surprising that our formulas do not give quantitative agreement with the experimental results. However, we do not expect that the qualitative features of our results could be changed in a more rigorous theory. The factor A makes the gap between theory and experiment narrower than before, but the factor S makes it broader. If we could include the effect of intervalley scattering, we would get another modulation factor I in the Eq. (2). This factor is probably greater than unity because the higher-energy electrons should lose energy through the intervalley scattering.

OUTLINE OF CALCULATION AND RESULT

Now we briefly outline the calculation and the results upon which the above discussion is based.

Including the effect of the shear waves, the deformation potential U for a spheroid becomes, as pointed out by Herring,[11]

$$U=E_d\{(1+\delta)\epsilon_{11}+\epsilon_{22}+\epsilon_{33}\} \tag{3}$$

where ϵ_{11}, ϵ_{22}, and ϵ_{33} are the diagonal components of the strain tensor along the principal axes of the spheroid. The subscript 1 is for the rotational axis, and 2 and 3 are for the other principal axes. $E_d\delta\epsilon_{11}$ is the shear wave component of the deformation potential. Owing to this component, we must modify the scattering matrix element for the longitudinal wave by a factor

$$1+\delta\cos^2\theta, \tag{4}$$

and also introduce a new matrix element for the transverse wave which can be obtained by multiplying

$$\delta\cos\theta\sin\theta \tag{5}$$

into the usual longitudinal matrix element of pure dilation interaction. Of course we must not forget to replace the longitudinal velocity C_l by the transverse velocity C_t in the expression of the latter matrix element. Here θ is the angle between the wave vector of the phonon and the rotational axis of the spheroid.

Using these matrix elements, we solve the Bloch equation by the same procedure as that used by Yamashita and Watanabe[7] or Davidow.[4] Three following assumptions (a), (b), and (c) are used:

$$\text{(a)}\qquad f(\mathbf{K},\mathbf{F},\mathfrak{M})=f_0(W,\mathbf{F},\mathfrak{M})+G(W,\mathbf{F},\mathfrak{M})\cdot\nabla_k W, \tag{6}$$

$$\text{(b)}\qquad \begin{cases} f_0(W\pm h\nu)=f_0(W)\pm h\nu df_0/dW \\ \qquad\qquad +\frac{1}{2}(h\nu)^2d^2f_0/dW^2, \\ \mathbf{G}(W\pm h\nu)=\mathbf{G}(W), \end{cases} \tag{7}$$

$$\text{(c)}\qquad h\nu\ll kT, \tag{8}$$

where $h\nu$ is the phonon energy. In Eq. (6) we assume that f_0 and $\mathbf{G}$ are functions of the wave vector $\mathbf{K}$ only through energy. Unfortunately this assumption leads to different expressions for the collision term along the three principal axes of the spheroid. Instead of doing a more elaborate calculation taking into account this anisotropy in the scattering, we shall simply use the average of these three collision terms. Also in the calculation of the drift term we use the approximation

$$\tfrac{1}{2}m_1v_1^2=\tfrac{1}{2}m_2v_2^2=\tfrac{1}{2}m_2v_2^2=\tfrac{1}{3}W. \tag{9}$$

Finally we get an expression for the current density in a strong field of electrons whose energy surfaces is a spheroid characterized by the mass tensor $\mathfrak{M}$.

$$\mathbf{J}=\frac{Ne\mu\pi}{2^{5/4}\Gamma(3/4)}\left(\frac{3}{\mathrm{Trace}\mathfrak{M}^{-1}}\right)^{\frac{1}{2}}\mathfrak{M}^{-\frac{1}{2}}\cdot\mathbf{F}. \tag{10}$$

TABLE I. $\sqrt{A_1}$.

Case	The direction of the current vector (100)	(110)	(111)
(A)	$\frac{1}{3}(2+r^{3/2})$	$\frac{1}{3}[1+2[\frac{1}{2}(r+1)]^{3/2}]$	$[\frac{1}{3}(r+2)]^{3/2}$
(B)	$\frac{1}{3}[1+2[\frac{1}{2}(r+1)]^{3/2}]$	$\frac{1}{3}[1+r^{3/2}+4[\frac{1}{2}(r+3)]^{3/2}]$	$\frac{1}{3}[1+[\frac{1}{2}(2r+1)]^{3/2}]$
(C)	$[\frac{1}{3}(r+2)]^{3/2}$	$\frac{1}{3}[1+[\frac{1}{2}(2r+1)]^{3/2}]$	$\frac{1}{3}[r^{3/2}+3[(r+8)/9]^{3/2}]$

[11] M. Shibuya, Phys. Rev. **95**, 1385 (1954).

Here μ is the weak field conductivity mobility defined by

$$\mu=\frac{2(2\pi)^{\frac{1}{2}}e\hbar^4MC_l^2}{3E_d^2(kT)^{\frac{1}{2}}\sqrt{\Delta}}\left(\frac{\mathrm{Trace}\mathfrak{M}^{-1}}{3}\right)\frac{1}{(1+2b\delta+c\delta^2)}, \quad (11)$$

and p is defined by

$$p=\frac{3\pi}{16}\left(\frac{\mu}{C_l}\right)^2\left(\frac{1+2b\delta+c\delta^2}{1+2a\delta+a\delta^2}\right)\left(\frac{3}{\mathrm{Trace}\mathfrak{M}^{-1}}\right)^2 \times\left(\frac{3}{\mathrm{Trace}\mathfrak{M}}\right)F\cdot\mathfrak{M}^{-1}\cdot F. \quad (12)$$

Here Δ is the determinant of the mass tensor and the three constants a, b, and c are functions of r and C given by

$$a=1/(2r+1), \quad (13)$$

$$b=\frac{1}{3}\left[\frac{2(\sqrt{r}+3)}{3(\sqrt{r}+1)^3}+\frac{2r^2\ln r}{(r-1)^3}-\frac{(3r-1)}{(r-1)^2}\right], \quad (14)$$

$$c=Cb+\frac{(1-C)}{3}\left[\frac{2}{(\sqrt{r}+1)^4}+\frac{(2r^2+5r-1)}{(r-1)^3}-\frac{6r^2\ln r}{(r-1)^4}\right]. \quad (15)$$

In order to discuss the hot-electron problem in actual crystals such as Ge and Si, we must consider the several spheroids in the Brillouin zone and sum up each current density (10) of these spheroids. Three special cases are considered: Case (A): 6 (or 3) spheroids along the cubic axes, Case (B): 12 spheroids along the face-diagonal axes, Case (C): 8 (or 4) spheroids along the body-diagonal axes.

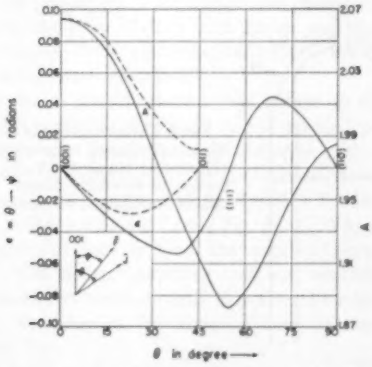


FIG. 2. A and ϵ *versus* θ in Case (C) with $r=0.06$. Here $\epsilon=\theta-\psi$; θ and ψ are the polar angles with respect to the Z-axis of current and field vectors respectively. For full curves these vectors are lying in the $X=Y$ plane and for broken curves they are lying in the $X=0$ plane.

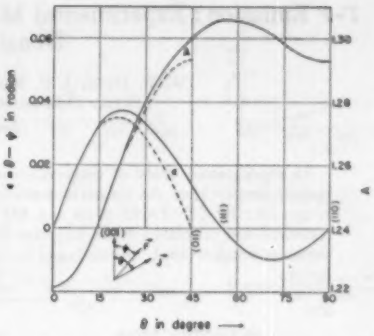


FIG. 3. A and ϵ *versus* θ in Case (A) with $r=0.2$.

The total current density J_T is given in each case by

$$\mathbf{J}_T=\frac{Ne\mu\pi}{2^{5/4}\Gamma(3/4)}\left(\frac{3}{\mathrm{Trace}\mathfrak{M}^{-1}}\right)^{\frac{1}{2}}\frac{1}{n}\sum_{i=1}^{n}p_i^{-\frac{1}{2}}\mathfrak{M}_i^{-1}\cdot\mathbf{F}, \quad (16)$$

where N is the total carrier density and n is the number of spheroids. Here p_i is a function of the electric field vector and of the direction of the rotational axis of the ith spheroid. Two factors A and S already introduced in Eq. (2) are derived from Eq. (16):

$$S=\left\{\frac{1+2a\delta+a\delta^2}{1+2b\delta+c\delta^2}\right\}^{\frac{1}{2}}, \quad (17)$$

$$A=A_1\frac{3}{(r+2)}\left(\frac{2r+1}{3r}\right)^{\frac{1}{2}}. \quad (18)$$

Here A_1 is dependent not only upon r but also upon both the spheroid arrangement and the direction of the current vector. In Table I, expressions for $\sqrt{A_1}$ are listed for the cases in which the current vector is directed along three symmetry axes of the cubic crystal. In Fig. 1, three S curves are plotted as a function of δ for $r=0.04$, 1 and 9, where we use 0.431 as the value of C. The anisotropy factor A and the deviation angle ϵ between the current and field vectors are the measures of the anisotropy in the hot-electron problem. They are plotted as functions of the direction of the current vector in Figs. 2 and 3, for n-type Ge and Si, respectively. In Fig. 2 we use the arrangement (C) with $r=0.2$. Two special cases are considered in these figures: For broken curves we let the current vector lie in the $X=0$ plane; the field vector then makes an angle ϵ with the current vector and lies in the same plane. For full curves both current and field vectors lie in the $X=Y$ plane.

ACKNOWLEDGMENTS

The author is much indebted to Professor J. Bardeen and to Dr. F. Ham for their kind discussions during this work, and to Dr. G. M. Hatoyama and Mr. W. Sasaki with whom this work was planned.

T-F Emission: Experimental Measurement of the Average Electron Current Density from Tungsten*

W. P. Dyke, J. P. Barbour, E. E. Martin, and J. K. Trolan
Physics Department, Linfield College, McMinnville, Oregon

(Received March 21, 1955)

An experimental study of temperature-and-field emission is described in which the average electron current density from the tungsten monocrystal was measured at several fields and temperatures in the ranges $2.5\times10^7<F<7\times10^7$ v/cm and $300°<T<2000°$K, respectively. The observed agreement between measured and calculated values supports the wave mechanical theory over larger combined ranges of the variables than previously reported, and further confirms that a high cathode temperature is an arc-initiating factor.

INTRODUCTION

A PREVIOUS paper[1] presented calculated values of the electron current density emitted from metals in the presence of both a high temperature T and a large surface electric field F. Under such conditions the emission process has been called "temperature-and-field emission" (abbreviated T–F emission), because current density was found to be sensitive to the values of both variables over their experimentally useful range. T–F emission is a process which is distinct from both field emission (large F and small T) and thermal emission (large T and small F).

This paper presents an experimental study of T–F emission in which the average electron current density from a tungsten monocrystal was measured at several fields and temperatures in the ranges $2.5\times10^7<F<7\times10^7$ v/cm and $300°<T<2000°$K, respectively. The experimental values are found to be in good agreement with those calculated by the methods of reference 1.

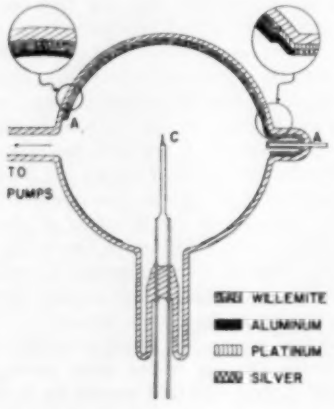


FIG. 1. Experimental electron projection tube: (*C*) needle shaped tungsten cathode; (*A*) aluminum-backed-willemite anode; anode-cathode spacing 4 cm. For details of anode construction see reference 7.

The increase of T–F emission current density with temperature, over the initial field current value, is consistent with the increase in current density observed just prior to the transition between field emission and the vacuum arc.[2] This result adds further support to the previous conclusion that a high cathode temperature is an arc-initiating factor.[1–4]

The present work extends the quantitative test of the wave mechanical emission theory to combined high values of temperature and field; previous quantitative tests at high fields have been restricted to low cathode temperatures.[2,5]

EXPERIMENTAL METHOD

For the present experiment, it was necessary simultaneously to measure or determine values of current density, electric field, work function and temperature at a metal cathode during the T–F emission of electrons from its surface. An experimental method was used which involved a straightforward modification of established field emission techniques,[2,3] the modification including primarily a means for maintaining and measuring a desired cathode temperature.

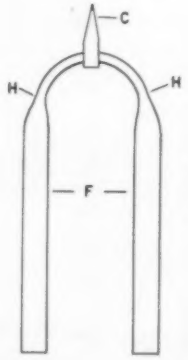


FIG. 2. Cathode assembly including: (*C*) needle shaped tungsten cathode; (*F*) support filament; (*H–H*), heated portion of support filament. For details of construction see reference 8.

* This work was sponsored by the U. S. Office of Naval Research.

[1] W. W. Dolan and W. P. Dyke, Phys. Rev. **95**, 327 (1954).

[2] Dyke, Trolan, Martin, and Barbour, Phys. Rev. **91**, 1043 (1953).

[3] W. P. Dyke and J. K. Trolan, Phys. Rev. **89**, 799 (1953).

[4] Dolan, Dyke, and Trolan, Phys. Rev. **91**, 1054 (1953).

[5] R. H. Haefer, Z. Physik **116**, 604 (1940).

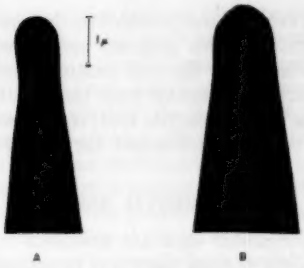


FIG. 3. Electron micrographs of needle shaped tungsten cathodes: (A), the cathode from which the data in Fig. 4 were obtained; (B) a "built-up" cathode which resulted from a 2-second application of an electric field of 3.4×10^7 v/cm and a temperature of 1900°K.

Electron emission was obtained from a tungsten single crystal in an electron projection tube.[6] With the aid of the observed emission pattern it was possible to detect and maintain conditions to be described under which the cathode surface was clean and free from geometric distortion. The work function for the clean tungsten surface is well known; the value used here, which is the average over all crystal faces, is $\phi = 4.5$ ev. The projection tube is shown in Fig. 1; it consists of a needle-shaped tungsten cathode C mounted on a support filament, and an aluminum-backed willemite anode A on which the emission pattern was viewed. Fabrication of the particular tube design in Fig. 1, has been described elsewhere.[7] The tube was sealed off at an ion gauge pressure of 10^{-9} mm of Hg and further evacuated by the use of tantalum getters (not shown), resulting in a residual pressure of 10^{-12} mm of Hg for chemically active gases. The evacuation procedure has been described previously.[8] Figure 2 shows an enlarged schematic drawing of the cathode C and support filament F. The cathode was tungsten, chosen for its mechanical strength, high melting point, and favorable electrical properties. That choice also permitted correlation of the present data with earlier experiments in field emission from cold tungsten. Fabrication of the cathode has been described.[8]

In order to provide electric field values up to 7×10^7 v/cm with reasonable values of the applied potential and electrode spacing, a needle-shaped cathode was used. The tungsten single crystal cathode approximated a cone of vertex angle 16 degrees with a hemispherical tip of radius $r = 5\times10^{-5}$ cm [Fig. 3(A)]. The cathode projected 0.4 mm beyond the support filament F, Fig. 2. The electron-emitting area was restricted to a portion of the hemispherical tip; a method for calculating the value of the emitting area has been discussed.[8]

[6] E. W. Mueller, Z. Physik **106**, 541 (1937).
[7] Dyke, Trolan, Dolan, and Grundhauser, J. Appl. Phys. **25**, 106 (1954).
[8] Dyke, Trolan, Dolan, and Barnes, J. Appl. Phys. **24**, 570 (1953).

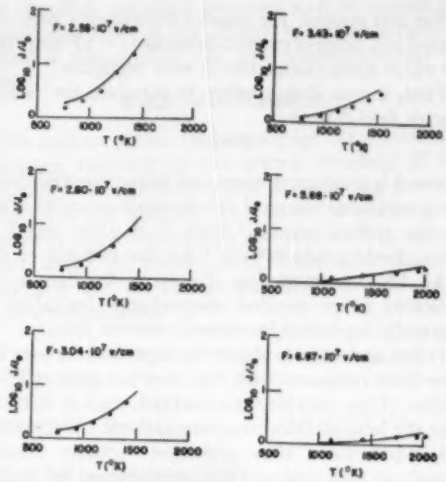


FIG. 4. Graphs relating $\log_{10}J/J_0$ to temperature T in °K for several values of the electric field. J_0 is the average value of the field current density from the cold cathode; J is the average value of the T–F current density from the heated cathode at the same field. Circles represent experimental data; solid curves show values calculated by the methods of reference 1.

Thermal energy was supplied to the emitter by conduction from the filament F (Fig. 2) which was maintained at a desired temperature by passing through it an electric current from an external source. Filament temperatures were measured in two ways; first, by use of a Leeds and Northrup optical pyrometer with a 1½-inch lens, which provided adequate optical resolution of the heated section of the filament HH, and second, by comparing measured and calculated values of thermal emission obtained from the same filament section. The cathode tip temperature was calculated from the measured filament temperature after correction for radiation loss by a method similar to that used by Mueller.[9]

The total electron emission current from the cathode, the applied voltage, the emission pattern photograph, and the cathode geometry were recorded by methods previously described.[2,3,7,8] From these data were calculated values for the cathode electric field[8] and the average cathode current density,[7] both of which will be useful in Fig. 4. Calculation of the latter by the indicated methods required that thermal emission from the emitter shank and support filament be negligible compared with T–F emission from the hemispherical emitting tip. This was achieved by restricting the temperature to values below 2000°K and minimizing the area of the heated portion of the cathode surface to the section HH, Fig. 2. For the latter purpose, the section HH was reduced from its original diameter of 10 mils to a diameter of 3 mils by electrolytic etch[8] in a concentrated solution of NaOH.

[9] E. W. Mueller, Z. Physik **126**, 642 (1949).

For two reasons, the present experiments were conducted in a range of current densities ($J<10^7$ amp/cm^2) for which space charge effects were negligible.[10]

First, it was then possible to calculate the cathode electric field F from

$$F=\beta V$$

where β is a constant computed from potential theory using measured values of the electrode geometrics, and V, the applied voltage.[8] With F in v/cm and V in volts, $\beta=3.9\pm0.2\times10^5$ cm^{-1} for the cathode of Fig. 3(A). Second, restriction of current densities to the indicated range avoided space-charge limitation of thermally dependent increases in current density.

It was necessary to obtain the experimental data in a time short compared with that in which geometric distortion of the cathode was observed, since it is known that the hemispherical tungsten cathode tip assumes a polyhedral shape after prolonged exposure simultaneously to high field and high temperature.[11,12] In that process, called "buildup," crystal faces with low indices such as the (110), (211), and (100) become planes of extended area and the intervening surfaces develop small radii of curvature. Buildup is illustrated in Fig. 3(B), a form which resulted during a time of 2 seconds at a field $F=3.4\times10^7$ v/cm and a temperature $T=1900$°K. Buildup undesirably intensifies electric field and current density and makes their measurement difficult at those cathode surfaces whose radii of curvature are decreased.

Buildup was avoided in the present work by use of two techniques. At the higher values of T and F, the electric field was applied for short time intervals by use of electronic pulse techniques; in this case the data were obtained during microsecond intervals on a single-pulse basis, both current and voltage being recorded as oscillographs by methods previously described.[2,3] At the lower fields, where the reduced current yield makes the use of pulse techniques difficult, the temperature was limited to a range in which buildup was negligible; in this case data were recorded in a few seconds using previously described direct-current techniques[3] which were modified only to include a current-metering system with fast response. For steady-state observations, current was determined from the voltage drop across a known resistance in series with the cathode, the voltage being measured by a Hewlett Packard vacuum-tube voltmeter.

Proof that buildup was negligible under the conditions used herein was provided by both emission pattern photographs and electron micrographs of the emitter which were obtained after the most severe usage; appreciable buildup is detectable both from changes in the emission pattern[11] and in the emitter geometry.[12] Additional evidence was provided by the reproducibility of the data. For this purpose, current and voltage were recorded from the cold emitter, then from the heated emitter and finally from the cold emitter, all at the same value of electric field; comparison of initial and final readings determined the reproducibility of the data.

EXPERIMENTAL RESULTS

The experimental data are presented in Fig. 4 in graphical form relating $\log(J/J_0)$ to cathode tip temperature T in °K for 6 values of the electric field in the range $2.58\times10^7<F<6.67\times10^7$ v/cm. J is the measured value of the T–F current density averaged over the emitting portion of the cathode hemispherical tip at the indicated temperature, while J_0 is the average value of the field current density from the cold cathode. The experimental data are represented by circles, and the solid curves represent values calculated by the methods of reference 1. While the absolute value of a particular electric field was known within 10%, the relative values of the various fields were known within 1%, which encouraged retention of the third significant figure in the field values used in the calculations on which Fig. 4 were based.

Corrections for the increase in work function with temperature,[13,14] and for the change in cathode electric field due to its geometric change during thermal expansion, were calculated for typical cases chosen from the present data. Since the corrections were comparable in magnitude to experimental error, they were ignored in preparing Fig. 4.

The experimental data for $F<4\times10^7$ v/cm were obtained by direct current techniques and are limited to temperatures less than 1550°K for which buildup was judged negligible. For fields $F>4\times10^7$ v/cm, the field was applied and data were obtained during single microsecond intervals, and no buildup was observed at the highest temperature shown, $T=2000$°K. As a check on the present work, measurements were made by a second method in which cathode current densities were determined by a photometric measurement of the light output from the phosphor screen.[7]

The experimental data quantitatively confirm the ratios of T–F and field emission current densities that are predicted in reference 1 for the stated ranges of temperatures and fields. The confirmation includes the predicted trend of decreasing temperature effect at higher fields. These results also appear to be in good agreement with the earlier work of Millikan and Eyring[15] over the more limited range of temperatures and fields reported in that work.

One result of the present investigation is further to establish a high cathode temperature as an initiating

[10] Barbour, Dolan, Trolan, Martin, and Dyke, Phys. Rev. **92**, 45 (1953).
[11] J. A. Becker, Bell System Tech. J. **30**, 907 (1951).
[12] R. O. Jenkins, *Reports on Progress in Physics* (The Physical Society, London, 1943), Vol. 9, p. 177.
[13] S. Seely, Phys. Rev. **59**, 75 (1941).
[14] J. A. Becker and W. H. Brattain, Phys. Rev. **45**, 694 (1934); A. R. Hutson, Phys. Rev. **98**, 889 (1955).
[15] R. A. Millikan and C. F. Eyring, Phys. Rev. **27**, 51 (1926).

factor for the transition between field emission and the vacuum arc.[1-4] The most direct experimental evidence to that effect was a large increase in current density over the initial field emission value which was observed just prior to arc initiation[2]; the increase can reasonably be attributed to the added effect of high temperature in view of the present results and an earlier calculation[4] of the expected resistive heating of the emitter.

A second result of the present investigation is further to support the validity of the wave mechanical theory which has been used to describe both the field emission[1,5] and the T–F emission[1] processes and to extend the quantitative test of the theory to higher combined values of temperature and field than previously reported.[3,5]

ACKNOWLEDGMENTS

The authors gratefully acknowledge the considerable assistance rendered by the several members of this laboratory. D. M. Barney and J. L. Boling provided fabrication and electron microscopy of emitters, respectively; tube construction was done by F. J. Grundhauser and F. M. Collins; L. M. Perry assisted in shop work and W. W. Dolan completed several mathematical calculations.

[1] A. Sommerfeld and H. Bethe, *Handbuch der Physik* (Verlag Julius Springer, Berlin, 1933), Vol. 24, Part 2, p. 441.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Hall Effect and Density of States in Germanium

E. M. Conwell
Physics Laboratory, Sylvania Electric Products, Inc., Bayside, New York
(Received April 20, 1955)

The usual assumptions that underlie the analysis of Hall data to obtain activation energy and density of states are based on the simple model of the band structure. They are re-examined here in the light of the known band structure of germanium. In particular, consideration is given to the limits of validity of the assumption that carrier concentration equals $1/R\epsilon$. Also, energy levels of electrons bound to donors are investigated and it is shown that the effect of excited states, particularly for Sb-doped germanium, cannot be neglected. The density-of-states mass values obtained from Hall data for p- and n-type material are considered. In the latter case, the results are shown to favor a band edge at the surface of the Brillouin zone rather than inside.

INTRODUCTION

ANALYSIS of variation of Hall constant with temperature has been an important source of information on activation energy of impurities and density of states in semiconductors. It has been customary to base such analysis[1] on two principal assumptions:

(1) The reciprocal of the product of Hall constant and charge on the electron equals total carrier concentration.

(2) Variation of carrier concentration, written for an n-type semiconductor, is given by

$$\frac{n(N_A+n)}{N_D-N_A-n}=\left(\frac{2\pi m^* kT}{h^2}\right)^{\frac{3}{2}} e^{(E_1-E_0)/kT}, \qquad (1)$$

where the notation is the usual one.[1]

These assumptions are known to be quite reasonable for semiconductors with the band structure usually assumed, the so-called simple model. The fact that neither the conduction band nor the valence band of germanium has this simple structure makes re-examination of these assumptions desirable for these cases. The first assumption, in particular, has been questioned for p-germanium because of the existence of two types of holes. In the next section the range of validity of this assumption will be examined. In the succeeding section the necessity for modifications in the equation which constitutes the second assumption will be considered. It will be shown in particular that the effect of excited states, which are not included in Eq. (1), will not be negligible in n-germanium. Corrections will therefore be required for activation energies and effective masses obtained from Hall data with the use of (1). In a final section, densities of states calculated from Hall data will be compared with theory. In the case of n-type material this is of particular interest because these results supplement the cyclotron resonance results and, at least in principle, permit determination of the number of valleys in the conduction band.

The considerations of this paper will be limited to high-resistivity samples, ones having carrier concentrations less than $10^{15}/\text{cm}^3$. One advantage of this is that complicating effects such as variation of activation energy with concentration and temperature should be small. Also, consideration will be limited to samples doped with impurities of small activation energy, of the order of 0.01 ev. For such samples the temperature

[1] See, for example, P. P. Debye and E. M.Conwell, Phys. Rev. 93, 693 (1954).

range of importance for determining the parameters in (1) is below 20°K, in which range cyclotron resonance results should be valid.

RELATION BETWEEN HALL CONSTANT AND CARRIER CONCENTRATION

Theory indicates that the proportionality factor between R and $1/ne$ is a function of the velocity-dependence of the relaxation time (in turn determined by the scattering mechanism) and of magnetic field intensity.[2] Its value is also dependent on the details of the band structure, so we must consider the cases of electrons and holes separately. In what follows we shall assume that only lattice scattering is operative, which should be reasonable for the samples of interest.

Consider first the case of electrons. Theoretical calculation of the proportionality factor between R and $1/ne$ have recently been carried out[3] for the actual conduction band structure. These yield the result that this factor is indeed equal to 1 within 10% for both low and high magnetic fields. The latter is significant because at the lower temperatures the low-field approximation should no longer be valid at the fields customarily used, a few thousand oersteds. Thus the first approximation mentioned in the introduction can be considered justified for n-type material.

The case of holes is more complicated because the hole population is made up of two types of holes with different mobilities. To treat this case we shall make the customary assumption that the mean free path for lattice scattering is independent of electron velocity.[4] This does not appear consistent with the observed temperature variation of the lattice mobility,[5] but should nevertheless yield order-of-magnitude results. If we denote heavy and light holes with subscripts 1 and 2, respectively, the usual low-field analysis for a two-carrier case leads to a proportionality factor:

$$R\Big/\frac{1}{(n_1+n_2)e}=\frac{3\pi}{8}\frac{(n_1\mu_1^2+n_2\mu_2^2)(n_1+n_2)}{(n_1\mu_1+n_2\mu_2)^2}. \tag{2}$$

The ratio n_1/n_2 should be approximately equal to $(m_1/m_2)^{\frac{3}{2}}$, which from cyclotron resonance results is 20/1.[6] The cyclotron resonance finding of approximately equal relaxation times for the two holes[6] leads to a mobility ratio $\mu_2/\mu_1=m_1/m_2=8/1$. Inserted in (2), these numerical values yield a low field proportionality factor of 2.5. Thus in the low-field region the approximation in question is quite poor for p-type material.

The situation is much better, fortunately, at high fields. It has been shown theoretically by Willardson *et al.* that at high fields the proportionality factor approaches unity for the case of a mean free path independent of velocity.[7] By similar methods this can be shown to be true for any velocity dependence of the mean free path. The field strength required for the high-field formulation to be valid within a few percent is determined by the stipulation that $[(9\pi/16)(\mu H)^2]$ be large, of the order of 25 or more. At 15°K, for example, the experimental value of $R\sigma$ for high-resistivity samples is about 2×10^5 cm^2/volt sec. With the values of n_1/n_2 and μ_1/μ_2 used previously, the criterion stated leads to the result that about 3000 oersteds or more are required for the heavy holes to reach the high field region. A field 1/64 as large will suffice for the light holes. The temperature range up to 15°K more than covers the region of constant slope in plots of $\log(1/Re)$ *vs* $1/T$ for high-resistivity p-type samples (see Fig. 1). This is the important region so far as obtaining values of activation energy and effective mass are concerned. Somewhat higher temperatures are required to reach the plateau corresponding to exhaustion. At these higher temperatures mobility is lower and fields higher than 3000 oersteds would be required to attain the high-field region. It is felt, however, that only a minor error in values of activation energy and effective mass will result from employing the first assumption of the introduction for a field of 3000 oersteds over the entire temperature range concerned.

EFFECTS OF EXCITED STATES

In a more general formulation than that of Eq. (1), it is useful to write:

$$\frac{n(N_A+n)}{N_D-N_A-n}=\frac{N_c}{F}, \tag{3}$$

where N_c is the "effective density of states in the conduction band" and F a factor determined by the impurity levels. To allow for any possible type of band structure, N_c is conveniently written:

$$N_c=2(2\pi m^{(N)}kT/h^2)^{\frac{3}{2}}, \tag{4}$$

where $m^{(N)}$ is the density-of-states mass introduced by Herring,[8] and essentially defined by (4). When excited states of the impurities are taken into account, the factor F takes the form

$$F=\sum_i g_i e^{-(E_i-E_c)/kT}, \tag{5}$$

where g_i and E_i are the statistical weight and energy, respectively, of the ith impurity level.[9]

For the samples we are considering here, the temperature range in which the excited states might have an effect on the parameters obtained from fitting the data

[2] See, for example, V. Johnson and K. Lark-Horovitz, Phys. Rev. **79**, 176 (1950).
[3] B. Abeles and S. Meiboom, Phys. Rev. **95**, 31 (1954).
[4] See, for example, W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950).
[5] A mean free path independent of electron velocity leads to $\mu_L \propto T^{-1.5}$, whereas the lattice mobility observed for p-germanium varies as $T^{-2.3}$.
[6] Dexter, Zeiger, and Lax, Phys. Rev. **95**, 557 (1954).
[7] Willardson, Harmon, and Beer, Phys. Rev. **96**, 1512 (1954).
[8] C. Herring, Phys. Rev. **96**, 1165 (1954).
[9] K. S. Shifrin, J. Tech. Phys. (U.S.S.R.) **14**, 43 (1944).

with (3) extends up to perhaps 14°K. When the hydrogen model of an electron bound to a donor (or a hole bound to an acceptor) is valid, the lowest excited state is so close to the conduction band that its contribution to F is less than 1% in this temperature range. In that case $F \simeq 2e^{-(E_1-E_0)/kT}$, and (3) is well approximated by (1).

Although it is known that the hydrogen model is not valid for a hole bound to an acceptor, there is evidence from p-silicon, which has the same type of band edge as p-germanium, that the spacing of the several lowest levels is approximately hydrogen-like[10] or even larger. Thus it appears reasonable to neglect all but the ground state in p-type material. In n-germanium, as we shall now show, there is good reason to believe that there are excited states very much closer to the ground state, and these cannot be neglected. According to cyclotron resonance results the minimum energy in the conduction band of germanium occurs for a $\mathbf{k}$ value in the [111] direction.[11] Because of symmetry requirements, if this $\mathbf{k}$ value is inside the Brillouin zone there are eight distinct minima, while if it is on the surface there are only four. Cyclotron resonance experiments do not distinguish between these two possibilities. In what follows when we are not referring specifically to one or the other of these cases we shall denote the number of minima by g_c.

In the approximation that the potential binding an electron to a donor ion is hydrogen-like, given by $-e/K\mathbf{r}$, it has been shown that the minima can be considered separately.[12] In this approximation each minimum contributes a wave function, φ_i, belonging to the same energy and the ground state of an electron bound to a donor is degenerate of degree g_c. (This neglects spin.) States above these g_c can be expected to be approximately as close to the conduction band as the excited states in the hydrogen model and will be neglected in what follows. Since the potential around the donor ion has only tetrahedral rather than spherical symmetry, the degeneracy mentioned is split up. To zero order, the ground-state wave functions are now given by

$$\sum_{i=1}^{g_c} \alpha_i \varphi_i,$$

where the α_i must obey the requirements of tetrahedral symmetry.[13] To obtain the correct linear combinations the g_c degenerate functions are used to set up a g_c-dimensional representation of the symmetry group of the system, T_d. Decomposition of this representation gives the irreducible representations to which the correct linear combinations must belong. In the case of four minima these are A_1 and T_1, which are one and three-dimensional,[14] respectively. The linear combination belonging to A_1 is the symmetric one, $\alpha_1=\alpha_2=\alpha_3=\alpha_4$. The three linear combinations belonging to T_1 are sufficiently characterized for our purposes by the fact that $\alpha_1+\alpha_2+\alpha_3+\alpha_4=0$ for all three. In the case of 8 minima, the 8-dimensional representation obtained decomposes into $2A_1+2T_1$. This simple doubling of the result obtained for $g_c=4$ results from the fact that no operation of T_d takes φ_i into the wave function for the opposite side of the same body diagonal. If the perturbing potential mixes wave functions from all eight valleys, as seems most likely, the two linear combinations belonging to A_1 consist of one which is the completely symmetric combination of all eight φ's and a second for which $\alpha_1=\alpha_2=\alpha_3=\alpha_4=-\alpha_5=-\alpha_6=-\alpha_7=-\alpha_8$.[15] The functions belonging to T_1 are such that $\sum_{i=1}^{8}\alpha_i=0$, as in the case of 4 valleys.

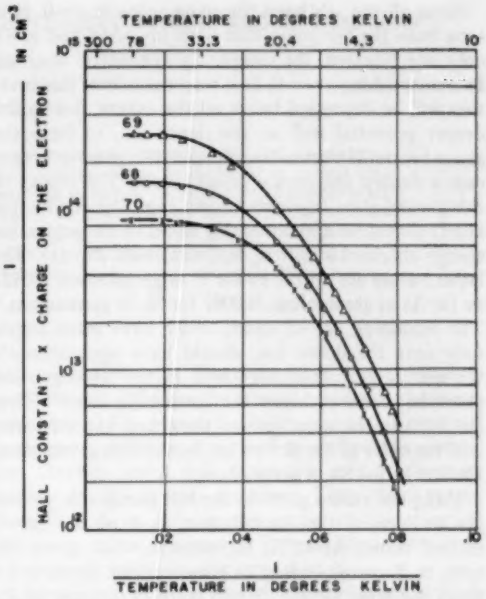


FIG. 1. Reciprocal of product of Hall constant and charge on the electron *vs* reciprocal temperature for some gallium-doped germanium samples. Magnetic field intensity was 3500 oersteds. The solid lines are the theoretical curves for an effective mass 0.3 m_0, an average activation energy of 0.011 ev. (From data of P. P. Debye.)

[10] Burstein, Bell, Davisson, and Lax, J. Phys. Chem. **57**, 849 (1953).

[11] Lax, Zeiger, Dexter, and Rosenblum, Phys. Rev. **93**, 1418 (1954).

[12] C. Kittel and A. H. Mitchell, Phys. Rev. **96**, 1488 (1954); M. A. Lampert, Phys. Rev. **97**, 352 (1955).

[13] The treatment followed here is that used for silicon by W. Kohn and J. M. Luttinger, Phys. Rev. **98**, 1178(A) (1955).

[14] See, for example, Eyring, Walter, and Kimball, *Quantum Chemistry* (John Wiley and Sons, Inc., New York, 1944), p. 388.

[15] If the perturbation did not mix φ_i's from opposite sides of the same body diagonal the linear combinations belonging to A_1 would be two symmetric sets of four φ_i's, each set having one φ_i from each of the four body diagonals. These would be complex conjugates of each other and have the same energy, resulting in a ground state of twice the degeneracy of the case considered in the text. The quantity N_c/F in (3) would then have the same value for four or eight minima, making it impossible to distinguish between the two cases.

Since all the φ's have the same value at $r=0$, it is seen from the foregoing that both for $g_c=4$ and $g_c=8$ only one function, the completely symmetric one, will be nonvanishing at $r=0$. It is reasonable that this function will be depressed below all the others, due to the deeper potential well at the donor ion, to form the ground state. If we now include spin, this gives us in each case a doubly degenerate ground state. The energy of this ground state (measured down from the conduction band) should be approximately equal to the activation energy obtained by fitting Hall data with Eq. (1). The latter values are 0.0120 ev for P in germanium, 0.0127 ev for As in germanium, 0.0097 for Sb in germanium.[16] The remaining $2g_c-2$ states, which have small amplitude near the donor ion, should have approximately the energy calculated for them in the hydrogen-like potential, 0.009 ev below the conduction band.[12] Thus the distance between ground state and excited states is of the order of 3×10^{-3} ev for As and P in germanium, 7×10^{-4} ev for Sb in germanium.

Using the values given in the last paragraph, we can get an idea of the contribution to F of the $2g_c-2$ excited states. At 12°K, for example, with $g_c=4$, the term in F corresponding to these excited states is 1.7 times the term for the ground state in the case of Sb, 0.08 times for As. Even at 4°K the excited state term is half that for the ground state in the case of Sb. It is evident that corrections to (1) due to excited states must be made for Sb in germanium. They are probably not serious for As or P in germanium.

It should be emphasized that if (1) is used instead of (3), (4), and (5) in fitting Hall data, the m^* value obtained may depend on the degeneracy and energies of the lower impurity states as well as on the conduction band states. For example, if the donor ground-state degeneracy of degree $2g_c$ had been unresolved, $F\simeq 2g_c e^{-(E_1-E_c)/kT}$, and the value of m^* obtained from (1) would be $m^{(N)}/g_c^{\frac{1}{2}}$. Only in the case where the ground state is doubly degenerate and the distance between the ground state and excited states is much greater than kT in the fitting range is the m^* value obtained equal to $m^{(N)}$.

DENSITIES OF STATES FROM HALL DATA

We should like now to consider the m^* values obtained from fitting Hall data under the assumptions of the introduction. It should be noted at the outset, however, that the results of such fitting are fairly insensitive to the m^* value. In good part this arises from the fact that the minority impurity concentration, which also occurs as a parameter, is unknown. Hall data taken at 3500 oersteds for some high-resistivity Ga-doped germanium samples[17] are shown in Fig. 1. Use of (1) with these data gave an average activation energy of 0.011 ev, in good agreement with the value of Geballe and Morin,[16] and a range of m^* values for acceptable fit of $0.2m_0$ to $0.6m_0$, where m_0 is the free electron mass. The theoretical curves obtained for $m^*=0.3m_0$ are shown in Fig. 1. It is expected in this case that the m^* value obtained from the use of (1) should equal $m^{(N)}$. For p-type germanium, cyclotron resonance results give:

$$m^{(N)}\simeq[(0.3)^{\frac{1}{2}}+(0.04)^{\frac{1}{2}}]^{\frac{1}{2}}m_0\simeq 0.3m_0.$$

Thus the density-of-states mass obtained from Hall measurements is at any rate consistent with the theoretical expectation for the valence band.

It is noteworthy that a question remains about the value of $m^{(N)}$ for holes at higher temperatures. In the 200°–300°K range, analyses of thermoelectric power data[18] and variation of Hall constant with temperature[7] have yielded considerably higher values than $0.3m_0$. The difference could conceivably result, for example, from a change in curvature of the bands. In both cases cited, however, it appears that the theoretical analysis must be revised. For example, in the latter case it has been assumed that the mean free path for lattice scattering is independent of electron velocity. As pointed out before, this does not appear consistent with the observed $T^{-2.3}$ variation of the lattice mobility. Further, repetition of the measurement of the microwave dielectric constant for p-germanium has yielded a value of $(0.3\pm0.05)m_0$ for this temperature range.[19] For the case of n-type samples, the considerations of the preceding sections indicate that use of (1) may not be too bad for the particular case of As-doped germanium. In this case excited states are probably far enough away from the ground state so that the m^* value obtained with (1) should be approximately equal to $m^{(N)}$. For the conduction band, theory indicates that $m^{(N)}$ is $g_c^{\frac{1}{2}}$ times the geometric mean of the principal masses for any one ellipsoidal surface of constant energy.[5] In germanium this mass[12] is $0.55m_0$ for $g_c=4$, $0.88m_0$ for $g_c=8$. Experimentally, it has been found that use of (1) with Hall data for As-doped germanium leads to m^* values in the range $0.2m_0$ to $0.5m_0$.[1] This favors $g_c=4$ rather than 8, or a band edge at the surface of the Brillouin zone. This evidence cannot be considered conclusive, however, until: (1) effects of excited states are considered more quantitatively, and (2) Hall measurements and fitting are repeated on samples of known minority impurity concentration, thereby eliminating most of the uncertainty in the m^* values.

It is a pleasure to acknowledge discussions with Dr. J. Birman and Dr. W. Kohn.

[16] T. H. Geballe and F. J. Morin, Phys. Rev. **95**, 1085 (1954).
[17] P. P. Debye, Phys. Rev. **91**, 208 (1953). For additional details on preparation and measurements see P. P. Debye and E. M. Conwell, reference 1.
[18] T. H. Geballe and G. W. Hull, Phys. Rev. **94**, 1134 (1954).
[19] J. M. Goldey and S. C. Brown, Phys. Rev. **98**, 1192(A) (1955).

Scintillation Response of Anthracene to Low-Energy Protons and Helium Ions*

E. J. ZIMMERMAN
University of Nebraska, Lincoln, Nebraska

(Received April 22, 1955)

Photomultiplier pulse amplitudes produced by light from anthracene crystals bombarded by protons and helium ions from 25 to 375 kev were measured. Both response curves (pulse height *vs* energy) are nonlinear, but that for protons bends away from the energy axis while that for helium ions bends towards it. Both curves extrapolate smoothly to the origin, indicating an absence of any nonscintillating surface layer more than about 10 kev thick. Present knowledge of stopping cross sections (dE/dx) for anthracene is inadequate to permit an unambiguous determination of the specific fluorescence (dL/dx) from the observed pulse amplitudes. However, some qualitative conclusions can be drawn concerning Birks' theory of the fluorescent process.

I. INTRODUCTION

THE emission of light from certain crystals bombarded by charged particles has been of considerable interest.[1-4] Organic crystals, such as anthracene, can be used to count individual slow ions if the response to such particles is known. Also, the light output, measured by the pulse amplitude produced in a photomultiplier tube, may give information about the luminescent process itself. The response of sodium iodide for slow heavy particles has been reported,[5,6] but all previous experiments with anthracene and other organic crystals have been done by slowing down particles with initial energies of several Mev. Because of straggling and uncertain range-energy relations, the low-energy response was poorly determined. In this experiment, ions accelerated by the Nebraska Cockcroft-Walton generator to accurately known energies between 25 and 375 kev were used to bombard anthracene crystals and the response curve has thus been extended below 100 kev.

II. EXPERIMENTAL METHOD

Figure 1 shows the experimental arrangement. The crystal chamber A was joined directly to the accelerator vacuum system. Protons, negative hydrogen ions, and singly charged helium ions were drawn from an rf ion source containing argon contaminated with a few percent of hydrogen and helium; this was necessary to reduce the intensity of the desired ion beam to a few hundred particles per second so that they could be counted individually. The energies of these ions were known within 2% from the generator voltage which was measured with a potential divider and potentiometer.[7] A magnetic analyzer was used to deflect the desired ions through about 40°, down the axis of the tube leading to the crystal chamber. They then passed through a $\frac{1}{16}$-in. hole in the center of the spherical aluminum mirror M, used to increase the efficiency of collection of light from the crystal at X.

Light from the crystal was transmitted to the cathode of the photomultiplier P through the Lucite window L, shaped to fit the curved surface of the phototube. Optical contact between L and P was made with Nujol mineral oil. In some runs, anthracene crystals about 1.0 cm×1.0 cm×0.4 cm were used; optical contact between these crystals and the window was also made with Nujol oil. Some thin (about 0.5 mm) crystals were also used. These were cemented to small glass blocks which were sealed to the window with oil. The aperture of the mirror and the diameter of the window were $1\frac{1}{4}$ in.; the effective solid angle for light from the crystal was about 3π.

The photomultiplier tube used was a special 17-stage tube developed by Schaetti,[8] and characterized by high gain, exceptionally low noise, and very high

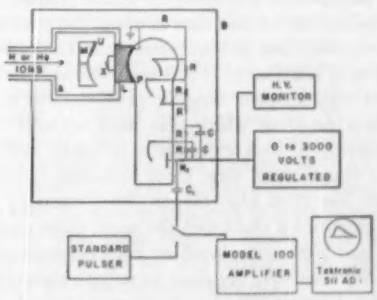


FIG. 1. Block diagram of apparatus.

* Supported in part by the U. S. Atomic Energy Commission and also in part by a Summer Grant by the Nebraska University Research Council.

[1] J. B. Birks, *Scintillation Counters* (McGraw-Hill Book Company, Inc., New York, 1953), and S. C. Curran, *Luminescence and the Scintillation Counter* (Academic Press, Inc., New York, 1953) contain extensive bibliographies.

[2] G. T. Wright, Phys. Rev. **96**, 569 (1954); J. Sci. Instr. **31**, 462 (1954); Phys. Rev. **91**, 1282 (1953).

[3] G. T. Wright and G. F. J. Garlick, Brit. J. Appl. Phys. **5**, 13 (1954).

[4] W. H. Robinson and W. Jentschke, Phys. Rev. **95**, 1412 (1954).

[5] S. K. Allison and H. Casson, Phys. Rev. **90**, 880 (1953).

[6] F. S. Eby and W. K. Jentschke, Phys. Rev. **96**, 911 (1954).

[7] Cook, Jones, and Jorgensen, Phys. Rev. **91**, 1417 (1953).

[8] N. Schaetti and W. Baumgartner. Z, angew. Math. u Phys. I. 268 (1950); Helv. Phys. Acta **23**, 524 and 869 (1950); Le Vide, No. 34-35, 1041 (1951).

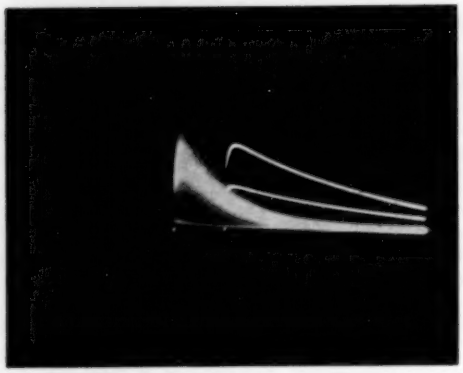


FIG. 2. Pulse distribution for 275-kev protons on anthracene. The sharp lines are reference pulses.

vacuum. The tube produces fewer than 2000 noise pulses per second, and these are almost certainly caused by single electrons from the cathode. The cathode quantum efficiency is about that for an RCA 5819. Since most of the pulses from ions were well above the single-electron pulses for incident energies of 75 kev or more, it would seem that the 5819 or similar tube could be used for efficient counting of these particles, the noise being biased off.

Dynode voltages were supplied by a resistance divider, the last three dynodes being by-passed directly to the high-voltage supply. About 2200 volts was sufficient to produce at the anode a mean pulse height of over one millivolt for the least energetic particles. The mean pulse height varied with high voltage in agreement with the gain *vs* voltage curves supplied by the manufacturer for this particular tube. Throughout any day's run, the high voltage was continuously monitored by comparing a fraction of it to a 67½-volt reference battery; the comparison was sufficiently sensitive to cause the mean pulse height to fluctuate less than 1% due to variation in tube voltage.

The pulses were amplified by a Los Alamos Model 100 amplifier with a rise time of about 0.6 microsecond and a clipping time of 8 microseconds. The pulses displayed on a Tektronix 511 AD oscilloscope were photographed with a Speed Graphic press camera giving a 4×5 inch negative which was used without further enlargement. A time exposure was taken sufficiently long to obtain more than three thousand pulses; this required less than one minute. Two standard pulses whose amplitudes bracketed the mean pulse amplitude of the ions were also placed on each photograph. The standard pulses were obtained by using a mercury relay (Western Electric 275A) to apply a measured voltage to an accurately-measured resistance attenuator; the relative sizes of the reference pulses were known to better than 2%. A sample photograph is shown in Fig. 2.

The mean pulse height was obtained by measuring the mean of the distribution of particle pulses, the height of the two standard pulses, interpolating, and converting to height in millivolts at the input of the amplifier. At the lower energies, where the pulse distribution becomes very broad, the determining of the "mean" pulse height by this method is questionable, since the judgment of the observer plays a significant role. However, above 100 kev, the mean amplitudes could be determined within 5 to 7%. The consistency between various measurements of the same photograph, and the consistency between various runs with different crystals was considerably better than this estimation, averaging about 3%.

III. EXPERIMENTAL RESULTS

Figure 3 summarizes the results of five runs with protons for four different anthracene crystals and of four runs with helium ions for three different crystals.† Each run normally gave a curve for both protons and helium ions with energies from 25 to 350 or 375 kev. Since the photomultiplier voltage could not be set exactly the same for each day's run, a single normalization point, chosen to be the smooth-curve pulse height for 250-kev protons, was used for each run. The bars shown on the experimental points are the average deviations from the mean of pulse height at a fixed

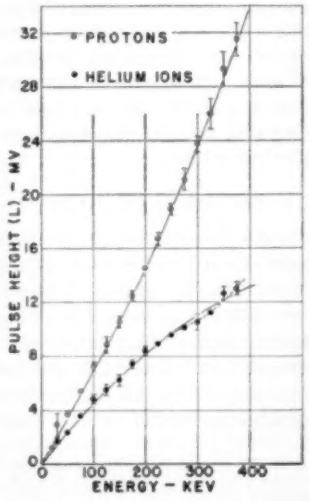


FIG. 3. Scintillation response of anthracene. ○ Protons. ● Helium ions. Ordinates are normalized pulse heights in millivolts at amplifier input. The solid curves are least-squares fitted to a quadratic form and are $L_p = 62E + 56E^2$ and $L_{He} = 50E - 42E^2$, with E in Mev.

† *Note added in proof.*—These results should be compared with the proton results of Fowler and Roos [Phys. Rev. **98**, 996, (1955), Fig. 2]. Their curve is somewhat more nonlinear, giving, for example, a ratio of 350-kev pulse height to 250-kev pulse height about 25% larger than obtained here. The cause of the difference is unknown.

energy after normalization; these average about 3% for both proton and helium ion curves. The various runs covered a time of several months, and the crystals were from two different batches supplied by the Harshaw Chemical Company.

In spite of the relatively high vapor pressure of anthracene, no contamination of the vacuum was observed. After several weeks *in vacuo*, the crystals developed a slight haze on the surface, but even this did not seem to affect the pulse-height distribution, which in every run taken, extrapolated smoothly to the origin. No fatigue or discoloration of the crystals was observed, but the integrated dose of heavy-particle bombardment was probably much too small to produce this effect.[1]

To be sure that the method was capable of producing correct results, a sodium iodide crystal was used and the response measured for protons and helium ions up to 250-kev energy. These data were in agreement with those obtained by Allison and Casson[5] for a similar experiment.

To facilitate comparisons with other experiments, the pulse height produced by the 624-kev internal conversion betas from Cs^{137} was measured, as well as that produced by the alphas from a natural uranium source having a mean energy of 4.3 Mev. On the scale used in Fig. 3, the 624-kev beta pulse height is 330 and the 4.3-Mev alpha pulse height is 230, both $\pm 7\%$. Using these data for normalization, the results of this experiment appear consistent with previous experiments at higher energies.[3,9]

Very rough estimates of the spread in pulse heights were made from the photographs. The percent resolution (full width at half-maximum divided by the mean pulse height) appeared consistent with the $E^{-0.7}$ variation found by Allison and Casson[5] for sodium iodide. For protons at 100 kev, the percent resolution was about 65%; it is estimated that at this energy the mean pulse size is about that which would be produced by pulses from the emission of four photoelectrons from the cathode.

IV. ANALYSIS AND DISCUSSION

Certain qualitative features of Fig. 3 are interesting. For protons, the curve of L (mean pulse height) *vs* energy is generally convex towards the energy axis, while that for helium ions is generally concave towards it. Thus the slope, dL/dE, for protons increases with energy, while that for helium ions decreases. Hence the light produced (dL) for a small energy loss (dE) in the crystal increases with the energy of the proton, but decreases with the energy of the helium ion. Since the maximum of the Bragg curve (specific ionization *vs* energy) for protons in anthracene is probably between 50 and 100 kev and for helium ions is probably between 250 and 400 kev, it is likely that, over most of the energy range observed in this experiment, the specific ionization for protons is decreasing while that for helium ions is increasing. Thus the observed curvature in the response curves is possibly correlated with the ionization density.

This is further substantiated by the possible inflection in the curve for helium ions, shown by the dotted line in Fig. 3. Not only does such an inflection point seem required by the data at higher energies,[6] but also the Bragg maximum for helium ions is near the energy of the possible inflection point in Fig. 3.

Since the response curves for both particles seem to continue in a very smooth way back to the origin, any surface effect tending to reduce the production of light is either very small or is a smooth function of the incident particle energy. An inactive surface layer thicker than about 10 kev seems definitely ruled out.

The response curves have generally been analyzed[1,4] by plotting the specific fluorescence, dL/dx, (x = range of particle) as a function of the differential energy loss with range, dE/dx. The slopes of the curves in Fig. 3 give values for dL/dE. If dE/dx is known, then dL/dx can be calculated. dE/dx can be found from the stopping cross sections, but unfortunately these have not been measured for the energy region below 500 kev for any solids other than metals[10] and ice.[11] By assuming the additivity of atomic stopping powers and assuming that the cross sections are independent of the physical state of the element, the stopping power of anthracene ($C_{14}H_{10}$) can be calculated from the work of Hirschfelder and Magee.[12] They computed, semiempirically, the stopping powers of several elements for energies as low as 5 kev, although the methods used may not be valid below several hundred kev. Using these values for dE/dx and combining them with the slopes from Fig. 3 (the dotted line was used for the helium ion curve) to get dL/dx, the results shown in Fig. 4 were obtained.

Figure 4 shows also the results of previous experiments taken from Robinson and Jentschke.[4] The linear response curve is obtained from electron data, not shown in this figure. The previous proton and alpha data are shown by the crosses. The alpha data show "saturation" of the specific fluorescence; that is, dL/dx becomes independent of dE/dx for large dE/dx. The proton dL/dx increases less than linearly with dE/dx, but there is no definite evidence of "saturation." (Although the ordinate of Fig. 4 is in "arbitrary units," normalization between various experiments is done in terms of the 624-kev beta pulse heights. Hence agreement of alpha and proton curves will indicate the degree of consistency between the experiments.)

Using the Hirschfelder-Magee dE/dx, the present

[9] Taylor, Jentschke, Remley, Eby, and Kruger, Phys. Rev. **84**, 1034 (1951).

[10] See a review article by S. K. Allison and S. D. Warshaw, Revs. Modern Phys. **25**, 779 (1953).

[11] W. A. Wenzel and W. Whaling, Phys. Rev. **87**, 499 (1952).

[12] J. Hirschfelder and J. L. Magee, Phys. Rev. **73**, 207 (1948).

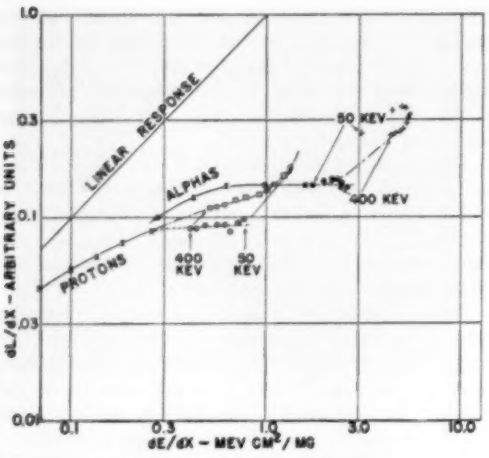


FIG. 4. Specific fluorescence vs differential energy loss for anthracene. Points × are taken from Robinson and Jentschke (reference 4). Points □ (protons) and + (helium ions) are derived using Hirschfelder-Magee stopping powers (reference 12). Points ○ (protons) and ● (helium ions) are derived from the "solids" stopping powers (see text). The "linear response" curve is determined by electron data, not shown here (reference 4).

proton data do lie on a smooth extension of the previous curve. However, two properties of this extension should be observed: (1) No evidence for saturation of specific fluorescence is seen at values for dE/dx much greater than values which produce saturation for alpha particles. (2) At very large dE/dx (low energies), dL/dx increases more than linearly with dE/dx. Such a situation is difficult to understand physically, although the theory of Chou[13] permits such a possibility. This was pointed out by Milton and Fraser,[14] who obtained a somewhat similar response curve for fission fragements in KI.

In a similar way, the present helium ion data also imply that (2) occurs for alpha particles, but only after the considerable region of saturation shown by the higher energy data. The up-turn of dL/dx must be rather abrupt, as shown by the dot-dash line in Fig. 4. Further, for the very high dE/dx obtained in this experiment, dL/dx appears to become a double-valued function of dE/dx. While this is possible physically, it is in a certain sense unsatisfactory.

It is possible that these results are produced by the use of incorrect values of dE/dx in anthracene. In an effort to obtain better values, the published data on the stopping powers of solids were reviewed,[10,11] and a set of stopping powers was derived on the following assumptions:

(1) Only experimental data for solids should be used (none of the data from experiments with gases was used).

(2) At each energy, the proton stopping cross section is a smooth function of atomic number.

(3) Atomic stopping cross sections for elements in a solid can be combined by addition to give the total cross section.

(4) The ratio of the alpha stopping cross section at $4E$ to the proton cross section at E is not four ($=Z^2$) but is that given by experiments on gold (the only solid for which comparisons have been made experimentally).

If the resulting cross sections are used with the slopes from Fig. 3, then the points labelled "Solids dE/dx" in Fig. 4 are obtained. These seem to fit with the high-energy data somewhat better than those obtained from the Hirschfelder-Magee dE/dx. Further, the saturation effect shows up consistently for both protons and alpha particles, and the present data for helium ions is in very good agreement with the observed saturation at lower values of dE/dx.

The saturation of specific fluorescence is expected on the basis of a simple theory proposed by Birks.[15] This gives

$$dL/dx=\phi(x/a_0)A(1+BdE/dx)^{-1}dE/dx, \qquad (1)$$

where A and B are constants, a_0 is a length characterizing the crystal, and ϕ is a function of x/a_0 which is 0.5 at zero range and which increases in a roughly exponential fashion to unity. From Fig. 4 and the previous experiments, it appears likely that in the present experiment for both protons and alpha particles, BdE/dx is sufficiently larger than unity that, approximately,

$$dL/dx=\phi A/B.$$

Since ϕ increases monotonically with energy, the present low-energy data should lie below the curves established by the high-energy experiments. Since, however, they actually lie above the previous data, or at most show saturation of specific fluorescence, ϕ cannot be varying greatly over this energy region. Hence a_0 must be rather small. Since the range of the most energetic protons used in this experiment is less than 6 microns in anthracene and that of the most energetic helium ions is less than 2 microns, it seems unlikely that a_0 can be as large as the 3 microns suggested by Birks.

A further conclusion on the basis of either value of dE/dx in Fig. 4 is that the constant B in Eq. (1) cannot be the same for both protons and alpha particles. (The constant A simply defines the scale for measuring pulse heights.) This substantiates the conclusion already reached by Robinson and Jentschke.[4]

It is, however, to be emphasized that an analysis of the present data according to Fig. 4 is not really possible until the stopping power, dE/dx, for protons and alphas in anthracene, or at least in some similar substance, is much better known than at present. It is most likely that neither set of points for the present data is correct, and Fig. 4 should not be taken too seriously. There is a great need for some accurate experimental measure-

[13] C. N. Chou, Phys. Rev. **87**, 903, 904 (1952).
[14] J. C. D. Milton and J. S. Fraser, Phys. Rev. **96**, 1508 (1954).
[15] See reference 1, p. 93.

ments of the stopping cross sections of solids at low energies.

V. ACKNOWLEDGMENTS

The author is indebted to the University of Nebraska Research Council for a Summer Fellowship Grant and to the University of Illinois Cyclotron Laboratory for the loan of the crystal chamber. Mr. C. E. Kuyatt and Mr. D. C. Lorents helped greatly in the experimental work. It is a very real pleasure to acknowledge the helpful discussions of this problem with Dr. F. S. Eby, Dr. W. K. Jentschke, and Dr. W. G. Cross.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Ferromagnetic Resonance in Two Nickel-Iron Ferrites

W. A. YAGER, J. K. GALT, AND F. R. MERRITT
Bell Telephone Laboratories, Murray Hill, New Jersey

(Received April 19, 1955)

Ferromagnetic resonance data are presented as a function of temperature from 4.2°K to 380°K in $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ and in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$. When the effects of eddy currents are corrected for, most of the energy dissipation in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ as measured by the resonance line width is shown to be due to a relaxation mechanism associated with the presence of divalent iron. The line width shows a maximum at 160°K at our frequency of measurement (24 000 Mc/sec). A thermodynamical analysis of the loss mechanism is presented which assumes that a torque associated with part of the free energy relaxes. This correlates the line width quite satisfactorily with the temperature dependence observed in the magneto-crystalline anisotropy and the spectroscopic splitting (g) factor.

I. INTRODUCTION

IN recent years, several studies of ferromagnetic resonance line width have been made in single crystals of ferrites.[1–5] It has, however, remained difficult to understand the mechanism of the losses in these materials. The purpose of the present paper is to report ferromagnetic resonance experiments on single crystals of two ferrite compositions which we feel throw some light on this mechanism. Quantitative chemical analysis for Ni and Fe together with the usual valences give $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ and $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ as the simplest formulas for the compositions on which we have worked. There is some possibility of small variations from these formulas in detail, but they are at least good approximations. The first of these compositions was grown by J. P. Remeika at Bell Laboratories using a slight variation on a method previously reported[6] which involves solutions of the constituent oxides in borax at high temperatures. The second composition was grown by Dr. G. W. Clark of the Linde Air Products Company by means of a flame fusion method.

Our data consist of: (1) resonant field as a function of crystal orientation, and (2) the width of the resonance line for both ferrites as a function of temperature from 4.2° to 380°K and as a function of crystal direction at several temperatures. Because they showed features of special interest, preliminary data on line width as a function of temperature in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ have been reported previously.[7]

From our data we are able to deduce for each temperature the magnetocrystalline anisotropy constants K_1 and K_2, the general damping parameter λ, and the spectroscopic splitting factor g. The analysis used to deduce these parameters is simple and has been presented elsewhere.[1,2,8] We will therefore merely summarize results which we must use. The technique of this experiment will also be discussed only briefly, since it is very similar to previous work done by us on small oriented spheres in a microwave cavity.[2,5]

We present in more detail an elementary thermodynamical theory of the line-width data in terms of the idea that a large part of the losses in ferrites arise from relaxations associated with electronic rearrangements like that which occurs in the low-temperature transition in Fe_3O_4. We suggest that the maximum in the line width at 160°K which we see in the data on $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ (see Fig. 3) is due to such a relaxation associated with a rearrangement of the valence electrons on the iron ions as the magnetization moves. The suggestion that motion of these valence electrons is involved in losses has been made before by

[1] L. R. Bickford, Jr., Phys. Rev. **78**, 449 (1950). Technical Report XXIII Laboratory for Insulation Research Massachusetts Institute of Technology October, 1949 (unpublished).
[2] Yager, Galt, Merritt, and Wood, Phys. Rev. **80**, 744 (1950).
[3] Galt, Yager, Remeika, and Merritt, Phys. Rev. **81**, 470 (1951).
[4] D. W. Healy, Jr., Phys. Rev. **86**, 1009 (1952).
[5] T. Okamura and Y. Kojima, Phys. Rev. **86**, 1040 (1952); T. Okamura, Sci. Rept. Research Inst. Tôhoku Univ. **A6**, 89 (1954).
[6] Galt, Matthias, and Remeika, Phys. Rev. **78**, 391 (1949). Remeika grew the $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ used in the present experiments with sodium metaborate as a flux, rather than the sodium tetraborate used in this reference.
[7] Galt, Yager, and Merritt, Phys. Rev. **93**, 1119 (1954).
[8] C. Kittel, Phys. Rev. **73**, 155 (1948).

Wijn and van der Heide for other ferrites at other frequencies.[9]

In terms of this theory of the line width, we are able to relate our results to the results of measurements of the viscous damping of an individual ferromagnetic domain wall as a function of temperature in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$. Such measurements have recently been made by one of us and discussed in these same terms.[10]

II. EXPERIMENTAL

The experimental technique was fundamentally the same as that previously described.[2] We will therefore only discuss those aspects of the later experiments which were different.

The most important changes were those required in order to reach low temperatures. The transmission-type cavity was at the end of a vertical run of the input and output wave guides which extended for some three feet. Of this vertical run about eight inches consisted of rectangular glass tubing of the same inside dimensions as the rest of the wave guide, with the inside walls silver plated to provide the necessary electrically conducting surfaces. The glass tubing abutted against suitable joints with the wave guide at each end, and the assembly was held together by wires which applied a compressive stress between the two sets of joints. This arrangement cut the heat leak down to quite acceptable values.

The cavity was placed inside a helium Dewar. The space between the sidewalls of this Dewar was connected through a stopcock to a pump lead so that a pressure of about 2 mm of N_2 gas at room temperature could be established in it before each run. When this Dewar contained liquid He, the N_2 condensed on the outside of the inner wall and minimized the diffusion of He through the glass. The Dewar extended well up above the glass sections in the wave guides. This helium Dewar, in turn, was inside another Dewar which could be filled with liquid nitrogen. A vacuum pump was connected to this Dewar assembly in such a way that it was possible to pump on the liquid helium.

These experiments were done with an electromagnet built by A. D. Little Company which could be rotated about a vertical axis. This made an important innovation possible in the measurement of magnetocrystalline anisotropy. The spherical single crystal of ferrite was mounted in the cavity in a fixed position with the (110) plane parallel to both the dc and the rf magnetic fields. The field required for resonance was observed as a function of crystal direction by rotating the electromagnet, not by rotating the crystal. This especially simplified the experiments at low temperatures.

The values of the applied magnetic field were measured by means of proton or lithium nuclear resonance as in previous work.[2]

The spherical samples were made with a device developed for the purpose by Bond.[11]

The cavity used in these experiments could not be tuned. Rather, the signal oscillator was tuned to the resonant frequency of the cavity. For this reason, line widths were measured with a method somewhat different from that used in reference 2. A sphere was placed loose in a polystyrene cup in the cavity, so that the dc field oriented it with a [111] direction parallel to the field. Then the field was rotated until it was so far from perpendicular to the rf field that the resonance line was quite weak; as a result the sample did not change the frequency of the cavity or the signal transmitted through significantly, even at resonance. Under these conditions, the dc magnetic field was modulated at 280 cps and the resulting *variation* in the transmitted signal was observed. If the conditions are as stated, this variation is proportional to the *slope* of the resonance absorption line [see Eqs. (6) and (16) in reference 2]. What we did experimentally, then, was to measure the separation in oersteds of the points of maximum slope on the resonance lines in this way. This has been converted to the corresponding Lorentzian line width at half-power points before plotting it in Figs. 3, 4, and 5. This technique could only be used with the dc magnetic field along a [111] crystal direction of course. When the field was along other crystal directions, line width data were taken with a sample fixed on a mounting rod. In these cases, the dc field was made to be far from perpendicular to the rf field by adjusting the rotational position of the mounting rod. The method was the same in all other respects.

III. ANALYSIS

Kittel[8] has derived the resonance condition taking account of sample shape, crystal anisotropy, and the spectroscopic splitting factor. He gives

$$\omega=\gamma H_{eff}, \tag{1}$$

where ω is the frequency in radians/sec, γ is the gyromagnetic ratio ($ge/2mc$), g is the spectroscopic splitting factor, and

$$H_{eff}=\{[H_z+(N_y+N_y'-N_z)M_z] \times[H_z+(N_z+N_z'-N_z)M_z]\}. \tag{2}$$

Here N_x, N_y, and N_z are the usual shape demagnetizing factors and N_z', N_y' are effective demagnetizing factors which are a convenient way of taking account of the crystalline anisotropy. In our experiments with spheres $N_x=N_y=N_z$, so these factors cancel out, and Eq. (2) becomes

$$H_{eff}=[(H_z+N_y'M_z)(H_z+N_z'M_z)]^{\frac{1}{2}}. \tag{3}$$

Since $N_y'M_z$ and $N_z'M_z$ turn out experimentally to be

[9] H. P. J. Wijn and H. van der Heide, Revs. Modern Phys. 25, 98 (1953). H. P. J. Wijn, thesis, Leiden, 1953 (unpublished). Separaat 2092, N.V. Philips Gloeilampenfabrieken, Eindhoven, Holland.

[10] J. K. Galt, Bell System Tech. J. 33, 1023 (1954).

[11] W. L. Bond, Rev. Sci. Instr. 22, 344 (1951).

at all times less than 10% of H_s at 24 000 Mc/sec, we may write:

$$H_{eff}=H_s+\tfrac{1}{2}(N_x'+N_x'')M_s. \quad (4)$$

N_x' and N_x'' are, of course, functions of crystal direction. They include terms in both K_1 and K_2 and we may therefore write:

$$H_{eff}=H_s+f_1(\theta)K_1/M_s+f_2(\theta)K_2/M_s, \quad (5)$$

where θ is the angle of rotation in a crystal plane. If we let θ be the angle in the (110) plane between H_s and the [100] direction, Bickford[1] has shown that

$$f_1(\theta)=2-(5/2)\sin^2\theta-(15/8)\sin^2 2\theta,$$
$$f_2(\theta)=\tfrac{1}{8}(\sin^4\theta-4\sin^2 2\theta+4\cos^4\theta)\sin^2\theta. \quad (6)$$

Equations (5) and (6) were used in conjunction with the measured values of H_s as a function of θ to determine H_{eff}, K_1/M_s, and K_2/M_s at each temperature of measurement. In most cases a least squares procedure was used to fit Eq. (5) to the experimental data.

The value of H_{eff} obtained from Eqs. (5) and (6) at each temperature is used in Eq. (1) to determine the g value for that temperature. The values of K_1 and K_2 are obtained from K_1/M_s and K_2/M_s by assuming that M_s is the saturation magnetization of the ferrite at the temperature of measurement.

Since our most accurate line width measurement is of the separation between points of maximum slope on the absorption line rather than between half-power points, we convert our data to the width of the corresponding Lorentzian line at half-power points. This is done by means of the following relation:

$$(2\Delta H)_{\text{max slope}}=(1/\sqrt{3})(2\Delta H)_{\text{half-power}}, \quad (7)$$

where $(2\Delta H)$ is the total width of the line in each case. All line widths presented in the present paper are $(2\Delta H)_{\text{half-power}}$.

It is convenient to derive the damping constant λ from the line width, since it is a more general parameter

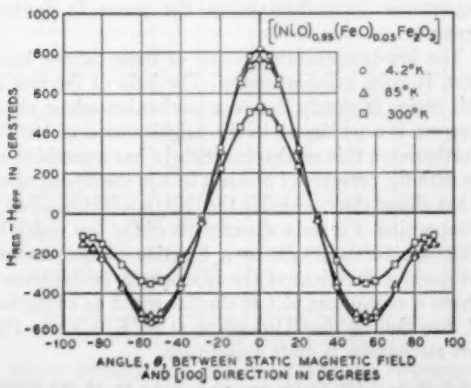


FIG. 1. $(H_{res}-H_{eff})$ as a function of angle between **M** and [100] direction in (110) plane for $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$.

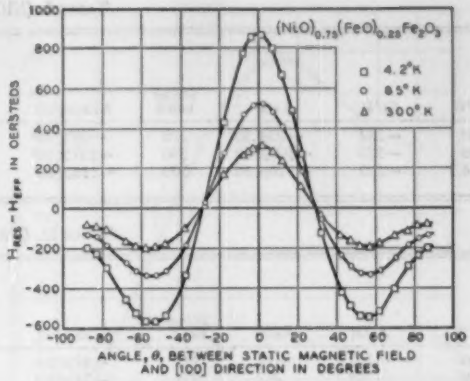


FIG. 2. $(H_{res}-H_{eff})$ as a function of angle between **M** and [100] direction in (110) plane for $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$.

with which to characterize the losses. λ is defined by writing for the equation of motion **M**:

$$d\mathbf{M}/dt=\gamma[\mathbf{M}\times\mathbf{H}]-(\lambda/M^2)[\mathbf{M}\times(\mathbf{M}\times\mathbf{H})]. \quad (8)$$

A relation between λ and the line width is given by Eq. (A-6) in reference 2. The relation given there is slightly incorrect, however.[12] It should be

$$(2\Delta H)_{\text{half-power}}/H_{eff}=2\lambda/\gamma M_s. \quad (9)$$

Note that if $(2\Delta H)$ is independent of crystal direction, Eq. (9) shows that λ is too. The α in reference 2 is equal to $\lambda/\gamma M_s$ in the notation of the present paper.

IV. RESULTS

Figures 1 and 2 show how the applied field required for resonance varies with crystal direction in the (110) plane for $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ and $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$, respectively.

It will be seen that our anisotropy results at 77°K for $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ differ substantially from those obtained by Healy on $NiOFe_2O_3$.[4] The differences are largest near the [100] direction. It seems doubtful that this disagreement can be accounted for by the deviation of the composition of our crystals from $NiOFe_2O_3$, since Healy's crystals came from the same source, and there is every reason to believe that they deviated from $NiOFe_2O_3$ in the same way. It should be noted that our data confirm the variation of K_1 with temperature as measured previously by a simpler but less accurate method at 60 cycles on stoichiometric $NiOFe_2O_3$.[6] We do not at present have a satisfactory explanation for this discrepancy.

Table I gives K_1/M, K_2/M, M, K_1, K_2, H_{eff}, ν, g, and λ for $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$. Table II gives the same data for $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$. Note that the values of K_1 in Table II are somewhat different from

[12] Equation (A-6) in reference 2 fails to take proper account of the dependence of l_x and l_y on H_s. This error occurs in reference 10 also.

TABLE I. $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$.

T°K	K_1/M	K_2/M	M (cgs units)	K_1(ergs/cc)	K_2(ergs/cc)	H_{eff}(oe)	ν(Mc/sec)	g(cgs units)	$\lambda(H$ parallel to [111] dir.) (cgs units) (corrected for eddy currents)
298	−254	−70±50	270	-68×10^6	$(-19\pm13)\times10^6$	7917	24 338	2.196	2.9×10^7
85	−370	−100±50	290	-107×10^6	$(-29\pm14)\times10^6$	7942	24 450	2.198	3.7×10^7
4.2	−392	−120±50	300	-118×10^6	$(-36\pm15)\times10^6$	7743	23 880	2.202	2.2×10^7

TABLE II. $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$.

T°K	K_1/M	K_2/M	M (cgs units)	K_1(ergs/cc)	K_2(ergs/cc)	H_{eff}(oe)	ν(Mc/sec)	g(cgs units)	$\lambda(H$ parallel to [111] dir.) (cgs units) (corrected for eddy currents)
300	−152	+16	322	-49×10^6	$+5.2\times10^6$	8195	24 388	2.13	10×10^7
159	−218	+3	338	-74×10^6	$+1.0\times10^6$	8094	24 461	2.16	14×10^7
135	−238	+14	339	-81×10^6	$+4.7\times10^6$	8054			14×10^7
85	−259	+24	340	-88×10^6	$+8.2\times10^6$	7980	24 471	2.19	10×10^7
4.2	−421	−15	342	-144×10^6	-5.1×10^6	7901	23 946	2.16	3×10^7

the preliminary data given in reference 10. The present data are the more accurate. Note also from Table II that the g factor in these single crystals is temperature dependent. This is not the case for $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$. This variation in g factor cannot be explained as a size effect; no such variation occurs in the value of the resonance field in the [111] direction, for example, among the 6, 7, and 9 mil spheres at any of the temperatures at which we took data.

The magnetizations of the $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ in Table I are obtained from Healy's[9] data and from measurements made in these Laboratories. The values of M_s for $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ in Table II are those used in connection with the domain wall studies on the same material made by one of us.[10] In this case the room temperature value is measured, and the values at other temperatures are determined from the assumption that M_s varies with temperature in this material in the same way as it does in Fe_3O_4 as measured by Weiss and Forrer.[12]

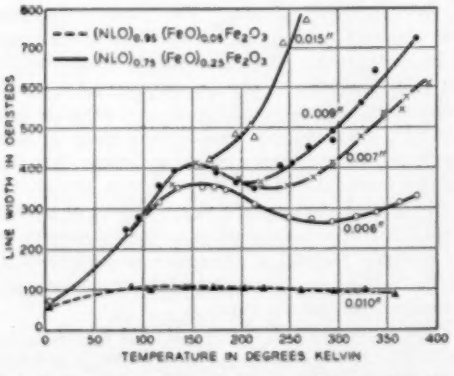


FIG. 3. Ferromagnetic resonance line width as a function of temperature for both ferrite compositions measured with dc magnetic field in [111] crystal direction. Note the size effect at high temperatures in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$; this is attributed to conductivity. The line is somewhat distorted in shape when this size effect is larger.

Figure 3 shows the line widths at each temperature for both compositions when the dc magnetic field is along the [111] direction. The values are plotted in the high temperature region for several sample sizes of the $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ composition. We believe that the rise in line width in this composition at these temperatures is due to the loss arising from the increasing conductivity of these crystals. This is confirmed by the fact that the size dependence in the line width characteristic of the relatively high conductivity ferrites[14] is observed in this region, and by the fact that crystals of the sort from which these spheres are cut have resistivities of only about 1 ohm-cm at room temperature.[15] The high-temperature end of the $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ curves may therefore be understood at least qualitatively in terms of a loss mechanism (eddy currents) which is not new and not of primary importance in understanding the losses in ferrites generally.

The low-temperature portion of these curves, however, is quite another matter. The bulk of the loss in this region is clearly due to a mechanism whose effectiveness is a maximum in the neighborhood of 160°K. Furthermore this mechanism clearly has something to do with the presence of divalent iron in the ferrite, since it has almost disappeared in the $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ composition. Figure 4 shows plots of the line width in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ as a function of temperature for various directions of the dc magnetic field. Figure 5 shows a continuous plot of the line width as a function of direction in the (110) plane at 85°K. Clearly this loss mechanism is quite anisotropic.

[12] P. Weiss and R. Forrer, Ann. phys. Series 10, 12, 279 (1929).
[14] Yager, Merritt, and Guillaud, Phys. Rev. 81, 477 (1951).
[15] T. H. Geballe and F. J. Morin (to be published).

The anisotropy in $(2\Delta H)$ for $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ is smaller than that for $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ as expected from the fact that $(2\Delta H)$ itself is smaller. Our measurements of line width in directions other than [111] for this material, however, are not sufficiently accurate to justify giving numerical values.

The absolute accuracy of the line width measurements gets worse as line intensity increases, for reasons discussed in Sec. II. We have made line intensity measurements which indicate that our relative errors, and our absolute errors in data taken at 85°K and below are 5% or less. The absolute errors in some of the line widths at temperatures above 85°K may be as large as 15%.

V. PHENOMENOLOGICAL THEORY

Having presented our data in the conventional way, we now turn to an elementary thermodynamical theory of the losses which explains at least qualitatively the maximum in the line width in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$. As we have seen above, it is fairly clear that this contribution to the losses arises from a relaxation associated with the fact that as the magnetization **M** moves the valence electrons on the iron ions are rearranged. The rearrangement does not occur fast enough to keep up with the motion of **M**, and as a result a net energy loss occurs.

It is not obvious, however, exactly what it is that relaxes, or to put it another way, what the time dependent interaction between the magnetic lattice and the electrons is. Some of the alternatives have been discussed in connection with the domain wall studies in reference 10. Following reference 10, we shall illustrate the mechanism by carrying out a calculation of the line width on the assumption that the torque on the magnetization associated with part of the free energy relaxes. As we shall see, however, we must assume that this part of the free energy must include another term in addition to the cubic anisotropy energy assumed in reference 10 in order to account for our experimental data on line width.

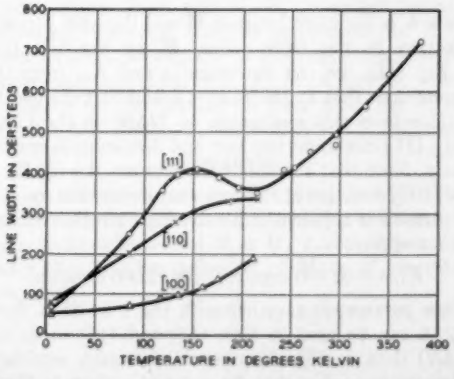


FIG. 4. Ferromagnetic resonance line width in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ as a function of temperature with the dc magnetic field in the three principal crystal directions. These data were obtained from the 0.009-in. sphere.

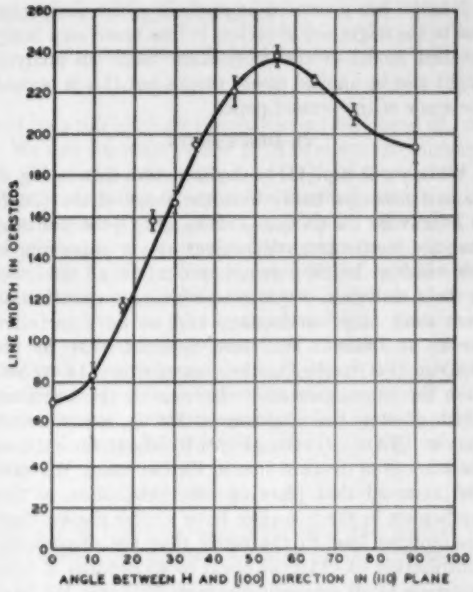


FIG. 5. Ferromagnetic resonance line width in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ as a function of angle between **M** and the [100] direction in the (110) plane. These data were taken at 85°K.

To proceed, we define our thermodynamical system as the whole magnetic lattice of the sample, which in our present experiments is one saturated magnetic domain. We assume as in reference 10 [Eq. (15)] that

$$\langle dw/dt\rangle_{\omega} = \langle dg/dt\rangle_{\omega}, \tag{10}$$

where the angular braces denote averages, dw is work done on the system, and dg is a change in the free energy of the system.

In the Appendix, the following expression for the line width at half-power points, $(2\Delta H)$, is derived on the basis of these thermodynamical considerations:

$$2\Delta H = \frac{1}{M_s}\left[\left(\frac{\partial^2 g_{1\omega}}{\partial\theta_x^2}\right)_{\theta_x=0} + \left(\frac{\partial^2 g_{1\omega}}{\partial\theta_y^2}\right)_{\theta_y=0}\right]\frac{\omega\tau}{1+(\omega\tau)^2}. \tag{11}$$

Here τ is the time associated with the rearrangement of the electrons, and $g_{1\omega}$ is the difference between the anisotropy energy when measured adiabatically (at high frequency compared to $1/\tau$) and when measured isothermally (at low frequency). This g should not be confused with the spectroscopic splitting factor. θ_x and θ_y are the angular deviations of **M** from equilibrium. It is clear that much more remains to be done in the theoretical analysis of the effect, but it is nevertheless instructive to compare the line-width data in Figs. 3 and 4 with Eq. (11). This is done in the next section.

Néel[16,17] has presented an analysis of the dissipation due to the migration of carbon in iron based on a fairly detailed model of the mechanism. Such an analysis might also be applied to our results but this is beyond the scope of the present paper.

VI. DISCUSSION

When we fit Eq. (11) to the line width data in Fig. 3, we first note (primarily from the shape of the $(2\Delta H)$ *vs* T curve for the $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ composition) that the mechanism which gives us our maximum in line width at 160°K does not account for all the losses in these materials. There is an additional contribution from some other mechanism, and we have corrected for it as follows. We have assumed that in the $(NiO)_{0.95}(FeO)_{0.05}Fe_2O_3$ the contribution to $(2\Delta H)$ from the rearrangement of electrons on the iron ions, which gives us the maximum at 160°K, is down from that in $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ by about the ratio of the amount of divalent iron in the two cases. We have also assumed that plots of the contribution of this mechanism in the two cases have similar shapes. Such assumptions lead to the result that the unexplained contribution to $(2\Delta H)$ is equal to 40 oersteds in both materials at all temperatures from 85°K up. We have therefore subtracted 40 oersteds from the line-width data given in Fig. 3 for $(NiO)_{0.75}(FeO)_{0.25}Fe_2O_3$ before fitting it to Eq. (11).

If we assume $\omega\tau=1$ at the temperature of the maximum in the line width (160°K), we find from the line width (see Fig. 3) at that point and Eq. (11) that $[(\partial^2 g_{1\infty}/\partial\theta_x^2)_{\theta_x=0}+(\partial^2 g_{1\infty}/\partial\theta_y^2)_{\theta_y=0}]=640M_s$. If we put this back into Eq. (11), and note that $\omega=1.54\times10^{11}$ in these experiments, we find that τ behaves as shown in Fig. 6 as a function of temperature. The accuracy of this plot, especially at 85°K, may be rather poor, but if we fit it to a function of the form,

$$\tau=\tau_\infty e^{\epsilon/kT}, \tag{12}$$

we find from the data in Fig. 6,

$$\epsilon=0.022 \text{ electron volts}, \quad \tau_\infty=1.0\times10^{-12} \text{ sec}.$$

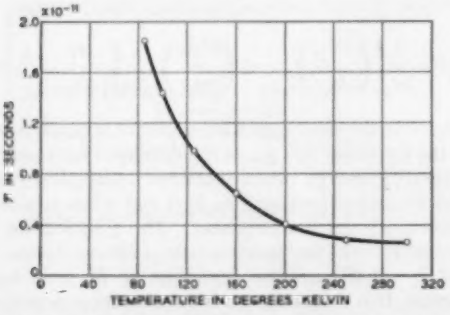


FIG. 6. Relaxation time τ as derived from the data in Figs. 3 and 4 corrected for eddy currents and Eq. (11).

[16] L. Néel, J. phys. radium **12**, 339 (1951); **13**, 249 (1952).
[17] J. K. Galt, Bell System Tech. J. **34**, 441 (1955).

It will be noted that this value of ϵ is in very good agreement with the activation energy for the production of carriers deduced from measurements of conductivity and thermoelectric power in a very similar ferrite crystal by Geballe and Morin.[18] Bloembergen and Wang[18] have also deduced a relaxation time which increases at low temperatures from ferromagnetic resonance data.

When we try to write an explicit form for $g_{1\infty}$ such that we can account for the line width data in Figs. 3, 4, and 5, we are forced to a more complicated form than previous thinking[10] led us to expect. A form which works is

$$g_{1\infty}=a(\theta_x^2+\theta_y^2)+K_{1\tau}(\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2). \tag{13}$$

The fact that the first term in Eq. (13) is not single valued in the angular position of **M** is at first unsettling. However, when we remember that the energy of the magnetization in a magnetic field is of the form $E=-MH(1+\cos\delta)$, we see that for small deviations δ of **M** from **H** the second-order term is the important one; it is $\frac{1}{4}MH\delta^2$, which is of the form of the first term in Eq. (13). In other words, this term implies a relaxation in the torque associated with part of the energy of the magnetization in the total effective field. The physical basis of this term is not entirely clear to us, but it leads us to expect a relaxation in the observed g factor. Inspection of Table II shows that our data confirm this expectation qualitatively. Table II also shows between 85°K and room temperature a variation in K_1 of the sort which the second term in Eq. (13) leads us to expect. The further change in K_1 observed between 4.2°K and 85°K must arise from other causes.

Now more quantitatively, if we differentiate Eq. (13), we find

$$\left(\frac{\partial^2 g_{1\infty}}{\partial\theta_x^2}\right)_{\theta_x=0}+\left(\frac{\partial^2 g_{1\infty}}{\partial\theta_y^2}\right)_{\theta_y=0}$$
$$=4(a+K_{1\tau})-20K_{1\tau}(\tfrac{1}{4}\sin^4\theta_0+\sin^2\theta_0\cos^2\theta_0), \tag{14}$$

where θ_0 is the angle between **M** and the [100] crystal direction in the (100) plane. If we use Eq. (14) in Eq. (11), we can determine a and $K_{1\tau}$ from the assumptions that at 160°K $\omega\tau=1$, and that the contributions from this mechanism to $(2\Delta H)$ in the [100] and [111] directions are zero and 320 oersteds respectively. Note that the difference between the $(2\Delta H)$ in the [100] direction at this temperature and the residual 40 oersteds is apparently due to eddy currents. We find for a and $K_{1\tau}$:

$$K_{1\tau}=-32\,000 \text{ ergs/cc}, \quad a=32\,000 \text{ ergs/cc}.$$

These parameters together with the τ at 85°K from Fig. 6 can be used in Eqs. (11) and (14) to fit the $(2\Delta H)$ data in Fig. 5 quite satisfactorily, especially when we remember that the value of τ given in Fig. 6 at 85°K may be in error by 25%. Furthermore the

[18] N. Bloembergen and S. Wang, Phys. Rev. **93**, 72 (1954).

data in Table II show that the change in K_1 observed between 300°K and 85°K is minus 39 000 ergs/cc. In view of the possibility that other factors may be causing a change in K_1 over this temperature range, we feel that this is in quite satisfactory *quantitative* agreement with the above value of K_{1r}. The remarks below Eq. (13) suggest that a may be represented by $(MH)_r/2$, so that the torque associated with an MH product of 64 000 ergs/cc is relaxing. This is approximately 2.5% of the value of (MH) for the saturation magnetization in the total effective field. As Table II shows, we observe a 3% change in g factor, which is how such an effect would be observed in our experiments. This also seems like quite satisfactory quantitative agreement, even though the physical basis of the first term in Eq. (13) is not clear.

Clogston[19] has analyzed this mechanism using a theory similar to that of Néel.[16] Because he assumes a much more detailed model than we do, he is able to derive the necessary terms in the free energy, rather than to assume them as is done in Eq. (13).

The λ we find from the present analysis [see Appendix Eq. (A11)] is very different in temperature dependence from that calculated from the domain wall data in reference 10. The difference, however, is easy to understand. It is due to the fact that the domain wall experiments are done in such a way that all angular frequencies associated with the motion are less than $1/\tau$. In that experiment we are therefore restricting ourselves at all temperatures to the region where $\omega\tau \ll 1$. In this case, as we see from Eq. (A11), $\lambda \sim \tau$; it therefore shows an exponential increase at low temperatures. In the present experiment, ω is held constant as a function of temperature, and $\omega\tau$ assumes values above one at low temperatures and below one at high temperatures. We therefore expect from Eq. (A11) a λ which shows a maximum in its variation with temperature, and this is what we observe.

The authors wish to express their gratitude to B. B. Cetlin, H. W. Dail, and M. R. Tiner for technical assistance. Chemical analyses were made by H. E. Johnson, J. F. Jensen, and J. P. Wright. Enlightening discussions were had with A. M. Clogston. Useful comments on the manuscript were made by R. M. Bozorth, J. F. Dillon, S. Geschwind, and S. Millman.

APPENDIX

If we take axes as usual with z along the equilibrium direction of the magnetization, and if we define θ_x as the angular deflection of $\mathbf{M}$ as M_x increases from zero, and θ_y the angular deflection of $\mathbf{M}$ as M_y increases from zero:

$$\frac{dg}{dt}=\frac{\partial g}{\partial \theta_x}\frac{d\theta_x}{dt}+\frac{\partial g}{\partial \theta_y}\frac{d\theta_y}{dt}, \qquad (A1)$$
$$=g_x'\frac{d\theta_x}{dt}+g_y'\frac{d\theta_y}{dt}.$$

Since $\mathbf{M}$ is precessing, we have,

$$\theta_x=\theta_{x0}\cos\omega t=(M_{x0}/M_z)\cos\omega t,$$
$$\theta_y=\theta_{y0}\cos\omega t=(M_{y0}/M_z)\cos(\omega t-\pi/2), \qquad (A2)$$

and from this of course it is easy to get $d\theta_x/dt$ and $d\theta_y/dt$.

We now calculate g_x' and g_y'. The electrons rearrange to minimize a free energy, so the form we use for g must be such that it is positive. Furthermore, we must form our analysis in such a way that a maximum is established for $\mathbf{g}'$, since even if the magnetization moves infinitely rapidly, $\mathbf{g}'$ does not become infinite. We therefore write g in the form $g=g_{1\infty}-g_1$, so that $\mathbf{g}'=\mathbf{g}_{1\infty}'-\mathbf{g}_1'$, where $\mathbf{g}_1'$ relaxes toward $\mathbf{g}_{1\infty}'$. Note that the torque relaxes downward from a value $\mathbf{g}_{1\infty}'$, associated with the adiabatic anisotropy energy (fast motions of the magnetization) to a value, zero, associated with the isothermal anisotropy energy (slow motions of the magnetization). We therefore assume that $\mathbf{g}_1'$ relaxes thus:

$$d\mathbf{g}_1'/dt=(\mathbf{g}_{1\infty}'-\mathbf{g}_1')/\tau, \qquad (A3)$$

where τ is the time associated with the process of rearranging the electrons. We have for the components of $\mathbf{g}_{1\infty}'$:

$$g_{1\infty x}'=\left(\frac{\partial g_{1\infty}}{\partial \theta_x}\right)_{\theta_x=0}+\left(\frac{\partial^2 g_{1\infty}}{\partial \theta_x^2}\right)_{\theta_x=0}\theta_x,$$
$$g_{1\infty y}'=\left(\frac{\partial g_{1\infty}}{\partial \theta_y}\right)_{\theta_y=0}+\left(\frac{\partial^2 g_{1\infty}}{\partial \theta_y^2}\right)_{\theta_y=0}\theta_y. \qquad (A4)$$

The terms in $(\partial g_{1\infty}/\partial\theta_x)$ and $(\partial g_{1\infty}/\partial\theta_y)$ are zero or cancelled by other torques when the magnetization is in equilibrium. It is therefore the terms in $(\partial^2 g_{1\infty}/\partial\theta_x^2)$ and $(\partial^2 g_{1\infty}/\partial\theta_y^2)$ which determine the variations in torque as the magnetization precesses and which are to be used in the component relaxation equations to determine g_{1x}' and g_{1y}'. The solutions are

$$g_{1x}'=\left(\frac{\partial^2 g_{1\infty}}{\partial \theta_x^2}\right)_{\theta_x=0}\frac{\theta_x\cos(\omega t+\varphi)}{[1+(\omega\tau)^2]^{\frac{1}{2}}},$$
$$g_{1y}'=\left(\frac{\partial^2 g_{1\infty}}{\partial \theta_y^2}\right)_{\theta_y=0}\frac{\theta_y\cos(\omega t+\varphi-\pi/2)}{[1+(\omega\tau)^2]^{\frac{1}{2}}}, \qquad (A5)$$

where $\tan\varphi=-\omega\tau$. We now have from (A1),

$$\frac{dg}{dt}=(g_{1\infty x}'-g_{1x}')\frac{d\theta_x}{dt}+(g_{1\infty y}'-g_{1y}')\frac{d\theta_y}{dt}. \qquad (A6)$$

The terms in $g_{1\infty x}'$ and $g_{1\infty y}'$ give zero when averaged over a cycle. When averaged, we find from A2 and A4,

$$\left\langle\frac{dg}{dt}\right\rangle_{\text{Av}}=\frac{1}{2}\left[M_{x0}^2\left(\frac{\partial^2 g_{1\infty}}{\partial \theta_x^2}\right)_{\theta_x=0}+M_{y0}^2\left(\frac{\partial^2 g_{1\infty}}{\partial \theta_y^2}\right)_{\theta_y=0}\right]\frac{1}{M_z^2}\frac{\omega^2\tau}{1+(\omega\tau)^2}. \qquad (A7)$$

Now the Landau-Lifshitz equation of motion for the

[19] A. M. Clogston, Bell System Tech. J. 34, 739 (1955).

magnetization is

$$d\mathbf{M}/dt=\gamma[\mathbf{M}\times\mathbf{H}]-(\lambda/M^2)[\mathbf{M}\times(\mathbf{M}\times\mathbf{H})]. \quad \text{(A8)}$$

From this we calculate

$$dw/dt=\mathbf{H}\cdot d\mathbf{M}/dt=(\lambda/M^2)(\mathbf{M}\times\mathbf{H})^2,$$

$$\left\langle\frac{dw}{dt}\right\rangle_{\omega}=\frac{\lambda H^2}{2M^2}(M_{x0}^2+M_{y0}^2). \quad \text{(A9)}$$

H here is the total effective magnetic field as given by Eq. (1) in the text. If we use this relation, and set $\langle dw/dt\rangle_{\omega}=\langle dg/dt\rangle_{\omega}$, we find from Eqs. (A8) and (A9),

$$\lambda=\left[\left(\frac{\partial g_{1\omega}^2}{\partial\theta_x^2}\right)_{\theta_x=0}+\frac{M_{y0}^2}{M_{x0}^2}\left(\frac{\partial g_{1\omega}^2}{\partial\theta_y^2}\right)_{\theta_y=0}\right]\times\frac{\gamma^2\tau}{1+(\omega\tau)^2}\frac{1}{1+(M_{y0}^2/M_{x0}^2)}. \quad \text{(A10)}$$

In our experiments most of the torque on the magnetization is due to the steady applied field, and is therefore circularly symmetrical about the z axis. Even the torque due to crystal anisotropy is circularly symmetrical in the [100] and [111] directions. We therefore assume that $M_{y0}^2\cong M_{x0}^2$, and Eq. (A10) becomes

$$\lambda=\frac{1}{2}\left[\left(\frac{\partial^2 g_{1\omega}}{\partial\theta_x^2}\right)_{\theta_x=0}+\left(\frac{\partial^2 g_{1\omega}}{\partial\theta_y^2}\right)_{\theta_y=0}\right]\frac{\gamma^2\tau}{1+(\omega\tau)^2}. \quad \text{(A11)}$$

Now by Eq. (9) in the text, we see that the width of the line at half-power points is $(2\Delta H)$:

$$2\Delta H=\frac{1}{M_s}\left[\left(\frac{\partial^2 g_{1\omega}}{\partial\theta_x^2}\right)_{\theta_x=0}+\left(\frac{\partial^2 g_{1\omega}}{\partial\theta_y^2}\right)_{\theta_y=0}\right]\frac{\omega\tau}{1+(\omega\tau)^2}. \quad \text{(A12)}$$

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Magnetostriction and Permeability of Magnetite and Cobalt-Substituted Magnetite*

L. R. Bickford, Jr.,† J. Pappis,‡ and J. L. Stull
State University of New York, College of Ceramics, at Alfred University, Alfred, New York

(Received March 28, 1955)

The principal magnetostriction constants of magnetite were determined over the temperature range from 120°K to 300°K by the strain gauge technique. Since λ_{111} and λ_{100} are different in sign and practically constant over this entire range ($\lambda_{111}\approx+80\times10^{-6}$; $\lambda_{100}\approx-20\times10^{-6}$) the change in direction of easy magnetization which occurs at 130°K is manifested by a change in structure of λ *vs* H curves in a polycrystalline sample. The substitution of small amounts of cobalt for divalent iron in polycrystalline magnetite causes a marked shift upwards in the temperature of the maximum in initial permeability, which in the case of magnetite occurs at 130°K. The temperature shift is practically linear with respect to cobalt ferrite content, the rate being *ca* 140°C/mole percent. Magnetostriction *vs* magnetic field curves for these specimens indicate that the shifted permeability peak is still associated with a change in direction of easy magnetization. The predicted anisotropy of cobalt ferrite, obtained by extrapolation of these results, is in reasonable agreement with the values measured directly by other investigators.

I. INTRODUCTION

THIS paper concerns the magnetostriction and magnetocrystalline anisotropy of magnetite over a temperature range in which its symmetry is cubic (above 120°K). The anisotropy is a rapidly varying function of temperature in the vicinity of 130°K, at which temperature the direction of easy magnetization changes from [111] to [100].[1] The temperature spectrum of initial permeability in this region, shown in Fig. 1, can be correlated directly with the absolute magnitude of the first-order anisotropy constant K_1 by the empirical relationship

$$\mu_0-1=39.1M_s^2/(|K_1|+1.21\times10^4), \quad (1)$$

in which μ_0 is the absolute magnitude of the initial permeability, and M_s is the saturation magnetization (obtained from Pauthenet's data[2]). The constant term in the denominator is presumably associated with magnetostrictive strain energy.

Evidence that the magnetostriction varies widely in magnitude and even changes sign at low temperature has been reported.[3,4] In order to check this temperature dependence directly, we have measured the principal

* This research was supported in part by the Office of Naval Research. Portions of this work were reported at two American Physical Society Meetings: Phys. Rev. **92**, 845(A) (1953); **94**, 1433(A) (1954).

† Now at International Business Machines Research Laboratories, Poughkeepsie, New York.

‡ Now at the Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts.

[1] L. R. Bickford, Jr., Phys. Rev. **78**, 449 (1950).

[2] R. Pauthenet, Ann. Physik **7**, 710 (1952).

[3] C. A. Domenicali, Phys. Rev. **78**, 458 (1950).

[4] J. L. Stull and L. R. Bickford, Jr., Phys. Rev. **92**, 845 (1953).

TABLE I. Saturation magnetostriction of magnetite single crystals at room temperature.

Type of crystal	Longitudinal Magnetostriction ($\times 10^6$) λ_{100}	λ_{110}	λ_{111}	Constants of 5-constant equation ($\times 10^6$) A_1	A_2	A_3	A_4	A_5
Natural	−16.7	65.4	90.6	−15.3	142.4	−9.43	−9.70	−50.5
Synthetic	−19.5	57.1	77.6	−24.5	123.3	−0.40	−4.76	−20.1
Synthetic (Calhoun)[a]	−19.4	98.7	86.4					

[a] B. A. Calhoun, Technical Report 68, Laboratory for Insulation Research, Massachusetts Institute of Technology, July, 1953 (unpublished).

magnetostrictive coefficients of magnetite single crystals from 120°K to room temperature.

During an investigation of the effect of minor additions of various foreign cations on the magnetite 120°K cubic to orthorhombic transition, we noticed that small amounts of cobalt profoundly affected the temperature of the permeability peak. We studied this effect by means of magnetostriction and permeability measurements.

II. EXPERIMENTAL PROCEDURE AND RESULTS

A. Magnetite Single Crystals

The constants of the familiar five-constant magnetostriction equation[5] between 120°K and 300°K were determined for two magnetite single crystals: a naturally occurring crystal from New York state and a synthetic crystal grown from the melt by Smiltens[6] and kindly furnished us by Professor A. R. von Hippel of Massachusetts Institute of Technology.

The samples were (110) plane circular disks *ca* 8 mm in diameter and *ca* 1 mm thick, with their edges rounded so that they approximated oblate ellipsoids. Resistance strain gauges (Baldwin type A19) cemented to the flat face parallel to the principal crystal directions ([001], [1$\bar{1}$0] and [1$\bar{1}$1]) were used to detect the deformation. Magnetostriction was measured as a function of crystal orientation with respect to an applied static magnetic field of *ca* 3500 oersteds (sufficient to achieve technical saturation) lying in the plane of the disk. The procedure outlined by Bozorth and Hamming[7] was used to obtain the magnetostriction constants, thus obviating the necessity of making any assumptions regarding the demagnetized state of the crystal. Values of saturation magnetostriction for the principal crystal directions were calculated from these constants.

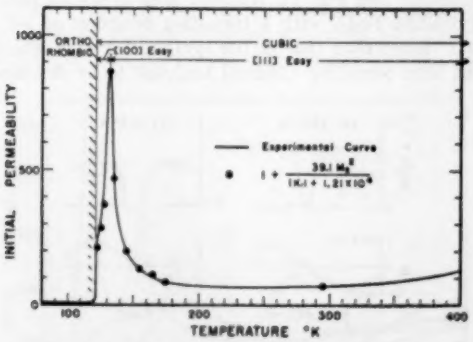


FIG. 1. Initial permeability of polycrystalline magnetite at 10 kc/sec as a function of temperature.

The strain-gauge manufacturer's room temperature value of gauge factor was assumed to be valid at the lower temperatures. It has been reported[8] that the calibration factor of this type of gauge does not change by more than 5% over the temperature range covered in this investigation. The probable error from all sources combined is estimated to be *ca*±20% at the lower temperatures, and somewhat less than this at room temperature.

Table I presents the five constants of the magnetostriction equation together with values of saturation longitudinal magnetostriction for the principal crystallographic directions of magnetite at room temperature. The temperature variation of λ_{100}, λ_{111}, and λ_{110} for the synthetic single crystal is shown in Fig. 2. Corresponding results for the natural crystal are similar.

B. Polycrystalline Magnetite

One of the interesting features of the magnetostrictive behavior of magnetite is that λ_{111} and λ_{110} are positive and larger in magnitude than λ_{100}, which is negative. As a result the net magnetostriction of polycrystalline magnetite is positive, whereas it is negative for most other ferrites. Since the relative magnitudes of λ_{111}, λ_{110}, and λ_{100} do not change appreciably between 120°K and 300°K, the structure of a curve of λ *vs* H in polycrystalline magnetite contains features which are characteristic of the direction of easy magnetization. Lack of knowledge concerning the exact details of the magnetization process in polycrystalline samples prevents us from making an exact analysis of a λ *vs* H curve. However, certain definite conclusions can be drawn from the qualitative behavior of such a curve in the vicinity of zero magnetization and in the region of large magnetic field.

A large part of the magnetostrictively active magneti-

[5] R. Becker and W. Döring, *Ferromagnetismus* (Verlag Julius Springer, Berlin, 1939), p. 275; see also reference 7.
[6] J. Smiltens, J. Chem. Phys. **20**, 990 (1952).
[7] R. M. Bozorth and R. W. Hamming, Phys. Rev. **89**, 865 (1953).
[8] E. E. Day and A. H. Sevand, Proc. Soc. Exptl. Stress Anal. **8**, 133 (1950).

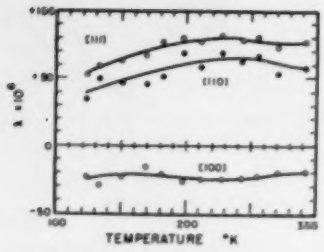


FIG. 2. λ_{100}, λ_{111}, and λ_{110} *versus* temperature for synthetic magnetite single crystal.

zation process at field strengths near the condition of zero net magnetization (demagnetized state) is domain wall motions (other than 180° domain walls) in those grains so oriented that one of the easy directions makes a small angle with the direction of the applied magnetic field. The magnetization process at high field strength is largely due to rotation of the magnetization in grains so oriented that one of the directions of most difficult magnetization makes a small angle with the field direction. Consequently the longitudinal magnetostriction of a polycrystalline material should have the same sign at weak fields as the longitudinal magnetostriction in the easy direction of the single crystal. In strong fields the λ *vs* H curve should approach its saturation value with a slope which has the same sign as the longitudinal magnetostriction in the hard direction of the single crystal. The features were observed in the λ *vs* H curves of polycrystalline magnetite.

The method used in preparing the specimen is described in the next section. The experimental arrangement, including sample size and shape, positioning of the strain gauge, and orientation of the magnetic field relative to the plane of the sample, was the same as for the single crystal measurements. However, for this experiment the sample remained stationary while the magnetic field intensity varied linearly and periodically so as to cause the sample to traverse a hysteresis loop. A special current source described by Levinson[9] supplied the magnet current. The output voltage of the strain gauge bridge was recorded as a function of time.

Figure 3 is a reproduction of two strain *vs* time curves for polycrystalline magnetite: one obtained at room temperature, where [111] is the direction of easy magnetization; and the other at 123°K, where [100] is easy. Curve (a) starts off from the state of zero magnetization with a positive magnetostriction (λ_{111} is positive, [111] easy, at room temperature) and approaches saturation with a negative slope (λ_{100} is negative). Curve (b), on the other hand, starts off from the demagnetized state in the negative direction (λ_{100} is negative, [100] easy), then reverses direction and approaches saturation with a positive slope (λ_{111} is positive). This type of reversal, called the Villari reversal, has been observed[10] to occur in polycrystalline iron, in which λ_{111} is negative and λ_{100} is positive. For magnetite the shape of the λ *vs* H curve changes from type (a) to type (b) at 130°K. Therefore the onset of a reversal in the curve serves as an indicator of the change in direction of easy magnetization from [111] to [100].

[9] J. Levinson, Phys. Rev. **95**, 651 (1954); Office of Naval Research Technical Report ONR III, Alfred University, January, 1954.

C. Cobalt-Substituted Magnetite

The Fe_3O_4 lattice will tolerate a deficiency of Fe^{2+} ions; that is, the spinel structure survives if the ratio Fe_2O_3/FeO exceeds unity ($Fe^{3+}/Fe^{2+}>2/1$). It has been demonstrated[11] that the 120°K cubic-orthorhomibc transition in magnetite is very sensitive to this type of deficiency. The temperature of the transition and magnitude of the discontinuity in electrical conductivity accompanying it both decrease with increasing ratio of Fe_2O_3/FeO.

One phase of our investigation was to study the behavior of the transition temperature in polycrystalline samples as the composition is changed by adding small amounts of foreign cations. It was necessary that the oxidation state remain stoichiometric with respect to the ratio of trivalent to divalent cations in order that the effects of the foreign cations would not be obscured.

The samples were prepared according to standard ceramic techniques, involving wet ball-milling of cp oxides and carbonates, dry-pressing, calcining at *ca* 1000°C, ball-milling, pressing into disks and firing at *ca* 1400°C. The calcining and firing operations were carried out in an atmosphere of CO and CO_2, in proportions varying with temperature according to the schedule reported by Smiltens.[6]

The optimum firing schedule was worked out with polycrystalline Fe_3O_4, using the temperature of the cubic-orthorhombic transition as quality criterion. Initial permeability measurements provided a very convenient means of monitoring the transition temperature, for this parameter drops precipitously at the transition (see Fig. 1). We were able to obtain polycrystalline Fe_3O_4 with a transition temperature only *ca* 1° lower than that of the synthetic crystal, which had been found by chemical analysis[6] to be stoichio-

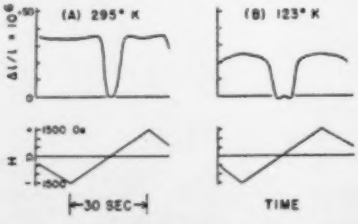


FIG. 3. Recorded magnetostriction *versus* time curves for polycrystalline magnetite. Corresponding magnetic fields are shown below. Curve (a) was obtained at room temperature, [111] easy direction; curve (b), at 123°K, [100] easy.

[10] M. Kornetzki, Z. Physik **87**, 560 (1933).
[11] E. J. W. Verwey and P. W. Haayman, Physica **8**, 979 (1941).

metric with respect to Fe^{3+}/Fe^{2+} ratio. Assuming that the substitution of small amounts of foreign cations into the magnetite lattice does not appreciably affect the oxidation equilibrium conditions, we used the same firing schedule for all samples, with slight variations in peak temperature. Cations added, in amounts not exceeding 5 mole percent, included the transition elements and Mg, Al, Cu, Zn, Ga and Ba.

Toroidal shaped specimens (*ca* 13 mm o.d.×10 mm i.d.×1 mm) were cut from the fired disks for permeability measurements. The edges of the remaining center core were rounded to form a sample for magnetostriction measurements. The toroids were served with two concentric monolayer windings covering most of the core (leaving a small section for contact with the copper-constantan thermocouple used for temperature measurements). The voltage induced in the inner winding by a sinusoidal 10-kc signal of known amplitude applied to the outer winding was measured. A simple calculation yielded the absolute magnitude of the permeability.

All cation substitutions were found to lower the transition temperature, with no significant correlation possible between the amount of lowering and the type of lattice site thought to be occupied by the ion. The amount of lowering per mole percent ranged from less than 1° (for barium) to 13° (for manganese). According to the work of Verwey and Haayman[11] the rate is approximately 9° per mole percent of holes in the octahedral cation lattice sites.

The substitution of foreign cations into the Fe_3O_4 lattice also caused a shift in temperature of the permeability peak shown in Fig. 1. Some cations raised the temperature of the maximum; others lowered it. In no case was the temperature interval between peak and transition decreased. With but two exceptions, manganese and cobalt, the amount of temperature shift per mole percent of foreign ion was less than 10°. It was very difficult to obtain good reproducibility in the properties of Mn-containing magnetite fabricated at different times. In some of the manganese-substituted samples the peak temperature was raised by as much as 20° per mole percent; in others it was lowered. The

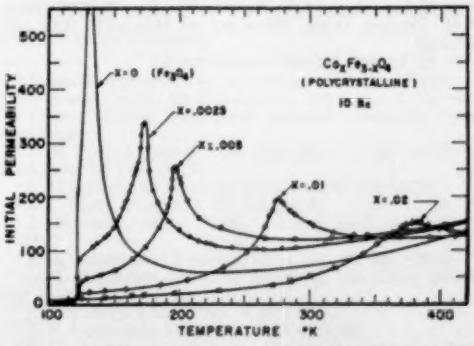


FIG. 4. Initial permeability of cobalt-substituted magnetite.

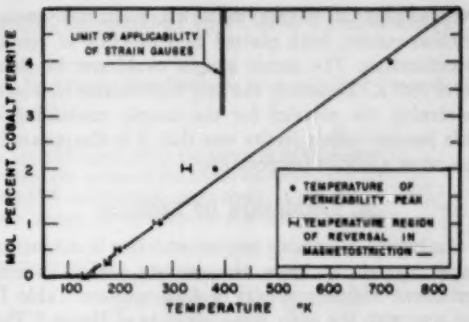

MOL PERCENT COBALT FERRITE
TEMPERATURE °K
LIMIT OF APPLICABILITY OF STRAIN GAUGES
TEMPERATURE OF PERMEABILITY PEAK
TEMPERATURE REGION OF REVERSAL IN MAGNETOSTRICTION

FIG. 5. Correlation between temperature of the permeability peak and temperature of first appearance of reversals (on cooling) in magnetostriction curves of cobalt-substituted magnetite.

addition of cobalt caused such a large and reproducible shift in the peak temperature that its effect can be regarded as unique among the cations studied.

Figure 4 shows permeability *vs* temperature curves for four different concentrations of cobalt, together with the curve for Fe_3O_4. Compositions are identified by the parameter X, which is the mole fraction of cobalt ferrite. The effect of the cobalt additions is to cause a very large shift upwards in the temperature of the permeability peak; also to broaden the peaks and decrease the value of permeability at the maximum. The amount of temperature shift is approximately linear with respect to cobalt ferrite content, the rate being *ca* 140° per mole percent.

Since the permeability peak in magnetite is a direct manifestation of the change in direction of easy magnetization, one would intuitively expect that the shifted peak in the cobalt substitution series could be attributed to the same cause. It was not possible to confirm this correlation directly, since single crystals of these compositions are not available. However, curves of longitudinal magnetostriction *versus* magnetic field obtained from the polycrystalline samples provided convincing indirect evidence that such a correlation is valid.

In the case of magnetite, since λ_{100} is negative and the average magnetostriction is positive, a low-field reversal occurs in the λ *vs* H curve of polycrystalline Fe_3O_4 when [100] is the preferred direction of the magnetization (Fig. 3). The presence of such a reversal in cobalt-substituted magnetite could be regarded as evidence that [100] is the direction of easy magnetization provided it could be assumed that the cobalt substitution does not affect the magnetostrictive characteristics of magnetite appreciably. To the extent that this assumption is valid the low-field magnetostriction reversals should first occur at the temperature at which the easy direction changes from [111] to [100].

Figure 5 shows the temperature of the permeability peak and the temperature range in which reversals

first appeared (on cooling) in the longitudinal magnetostriction curves; both plotted as a function of cobalt concentration. The strain gauges could not be used above 390°K. Therefore, the only information obtained concerning the reversal for the sample containing 4 mole percent cobalt ferrite was that if it disappears it does so at a higher temperature.

III. DISCUSSION OF RESULTS

Our results concerning magnetostriction in magnetite single crystals at room temperature are in general agreement with the results of Calhoun (see Table I) and also with the early measurements of Heaps.[12] The only noteworthy disagreement concerns the relative magnitudes of λ_{110} and λ_{111}.

The principal magnetostriction constants of Fe_3O_4 have been found to be relatively insensitive to temperature between 120°K and 300°K. It is probably for this reason that the temperature spectrum of initial permeability can be correlated very well with magnetocrystalline anisotropy, with the stress anisotropy expressed by a constant [Eq. (1)]. The apparent changes in sign of λ_{100} previously reported[3,4] to occur at low temperature were probably caused by spontaneous changes in the domain distribution of the demagnetized state accompanying the change in direction of easy magnetization. The previously reported[13] spontaneous deformation accompanying the change in direction of easy magnetization in the synthetic single crystal serves as evidence of a change in initial domain distribution.

Substitution of small amounts of cobalt into the magnetite lattice has been found to cause a large shift upwards in the temperature of the permeability peak (Fig. 4). The correlation evident in Fig. 5 between the temperature of the peak and the temperature at which reversals appeared in the magnetostriction curves is regarded as sound evidence that the permeability peak in all cases is associated with a change in direction of easy magnetization. The divergence of the two sets of data at the higher cobalt concentrations may well be due to a breakdown of the assumption that cobalt does not affect the magnetostrictive behavior of magnetite. The magnetostriction of cobalt ferrite is negative and very large.[14] Therefore, the addition of more than one percent of cobalt ferrite to magnetite might be expected to upset the qualitative relationship between magnitudes of the magnetostriction in the various crystal directions. We therefore conclude that the effect of the cobalt additions is to raise the upper limit of the temperature region in which the cube edge is the preferred orientation of the spontaneous magnetization.

If the straight line of Fig. 5 is extended to 100% cobalt ferrite, it falls at a value greatly in excess of the Curie temperature of cobalt ferrite. This extrapolation infers that [100] is the easy direction for cobalt ferrite over its entire cubic range. Moreover, it is possible to estimate from our data the value of the anisotropy constant of cobalt ferrite at room temperature. The anisotropy constant for Fe_3O_4 at room temperature is -1.12×10^5 ergs/cc.[1] We can assume that the constant for 1.18% cobalt ferrite is zero, since the permeability peak falls at room temperature for this composition. Making a linear extrapolation to higher percentages of cobalt ferrite we predict a value of $+7.5\times10^6$ ergs/cc for 80% and $+9.4\times10^6$ ergs/cc for pure cobalt ferrite. Recently the anisotropy constants of single crystals (approximately 80% cobalt ferrite) have been determined independently by two different investigators. Shenker[15] found that K_1 at room temperature is $+3.82\times10^6$ erg/cc; Bozorth[16] found a value of $+3.4\times10^6$ ergs/cc. In view of the extensive nature of our extrapolation the agreement between predicted and observed values of anisotropy constant is regarded as satisfactory.

All the observed effects are completely understandable, phenomenologically at least, on the basis of a very strong interaction between the spins of the cobalt ions and the lattice. This spin-lattice interaction aligns the cobalt spins parallel to a cube edge of the spinel lattice. The presence of a small amount of cobalt in the magnetite lattice is sufficient to align the entire spin system in this direction. Presumably cobalt would have the same effect on other ferrites.

[12] C. W. Heaps, Phys. Rev. **24**, 60 (1924).
[13] L. R. Bickford, Jr., Revs. Modern Phys. **25**, 75 (1953).
[14] R. M. Bozorth and J. G. Walker, Phys. Rev. **88**, 1209 (1952).
[15] H. Shenker, thesis, University of Maryland, 1955 (unpublished).
[16] R. M. Bozorth (private communication).

Specific Heat Anomaly of Ferroelectric Cadmium Niobate at the Curie Temperature*

H. R. DANNER† AND R. PEPINSKY
X-Ray and Crystal Analysis Laboratory, Department of Physics, The Pennsylvania State University, University Park, Pennsylvania

(Received May 9, 1955)

The anomalous specific heat of ferroelectric $Cd_2Nb_2O_7$ has been measured in the region of its Curie temperature, $T_c=193°K$, using a vacuum adiabatic calorimeter. The values of the transition energy and the transition entropy change are 18 ± 2 cal/mole and 0.09 ± 0.01 cal/°K-mole, respectively. A comparison between the experimental values of the transition entropy change, the spontaneous polarization, and the Curie constant is made.

INTRODUCTION

COOK and Jaffe[1] recently reported $Cd_2Nb_2O_7$ to be ferroelectric below 170°K. Subsequent dielectric measurements by other investigators[2,3] on ceramic samples of this compound, exhibiting Curie temperatures between 182°K and 188°K, have revealed the existence of a second phase transition at 85°K. Cadmium niobate, in addition to having similar chemical properties, contains metal octahedra, MO_6, characteristic of previously studied oxide ferroelectrics such as $BaTiO_3$. However, it differs from the latter in that its structure above the Curie point is face-centered cubic of the pyrochlore type.[4] An x-ray study of single crystals has shown that $Cd_2Nb_2O_7$ transforms to a tetragonal structure below the Curie point.[5]

Specific heat measurements have played an important role in the study of ferroelectrics, since a knowledge of the anomalous behavior of the specific heat enables one to calculate the change in energy, ΔE, and the change in entropy, ΔS, associated with the transition. These values, in turn, form an integral part of the experimental data needed in theoretical treatments of the phenomenon of ferroelectricity. In KH_2PO_4, for example, Stephenson's experimental value[6] for ΔS has given strong support to the Slater theory,[7] which is based on a statistical mechanical arrangement of the hydrogen bonds. In $BaTiO_3$, the experimental values of ΔE at the three transitions are in qualitative agreement with the values derived from the thermodynamic treatment of Devonshire.[8] Finally, as discussed later, ΔS is closely related to the spontaneous polarization and the Curie constant.[9]

APPARATUS AND METHOD

The specific heat measurements were made using a vacuum adiabatic calorimeter of the classical Nernst type. Numerous references concerning the essential details of construction and the method of measurement have been given in the literature.[10] Essentially, this apparatus allows one to add a measured amount of heat electrically to the sample, which is thermally insulated in an evacuated adiabatic shield system, and to observe the temperature rise produced. The specific heat of the sample is then computed from the equation

$$C_p=\left[\frac{VIt}{J(T_2-T_1)}-C_0\right]\frac{M}{m},$$

where V and I are the average voltage and current supplied for a time t, J is the mechanical equivalent of heat, T_1 and T_2 are the initial and final temperatures of the sample respectively, and C_0 is the heat capacity of the sample container. Multiplication by the ratio of the molecular weight M to the weight of the sample m gives C_p in cal/°K-mole. In practice the amount of heat added is small, so that the value of C_p may be taken as the value of the specific heat of the sample at the temperature $(T_1+T_2)/2$. A series of such heating periods yields a plot of C_p against T. In the present study, the temperatures were measured by observing, with a Leeds and Northrup type K-2 potentiometer, the thermal emf developed by a single copper-constantan thermocouple.

The operation of the apparatus was checked by measuring the anomalous specific heats of KH_2PO_4 and $BaTiO_3$, the latter at its lowest transition. For KH_2PO_4 a maximum in the specific heat curve occurred at 123.5°K, and the value of the transition energy, $\Delta E=\int\Delta C_p dT$, was 87 ± 6 cal/mole. The corresponding entropy change, $\Delta S=\int(\Delta C_p/T)dT$, was 0.72 ± 0.05 cal/°K-mole. These results are in excellent agreement with those reported by Stephenson.[6] For $BaTiO_3$ the maximum value of the specific heat appeared at 210.5°K, and the integrated values of ΔE and ΔS were 11 ± 2 cal/mole and 0.05 ± 0.01 cal/°K-mole

* Research supported by a contract with the Wright Air Development Center, Air Research and Development Command.
† Present address, Brookhaven National Laboratory, Upton, New York.
[1] W. R. Cook, Jr., and H. Jaffe, Phys. Rev. **88**, 1426 (1952); and Phys. Rev. **89**, 1297 (1953).
[2] G. Shirane and R. Pepinsky, Phys. Rev. **92**, 504 (1953).
[3] J. K. Hulm, Phys. Rev. **92**, 504 (1953).
[4] A. Bystroem, Arkiv. Kemi., Mineral. Geol. **A18**, No. 21 (1944).
[5] Jona, Shirane, and Pepinsky, Phys. Rev. **98**, 903 (1955).
[6] C. C. Stephenson and J. G. Hooley, J. Am. Chem. Soc. **66**, 1937 (1944).
[7] J. C. Slater, J. Chem. Phys. **9**, 16 (1951).
[8] A. F. Devonshire, Phil. Mag. **40**, 1040 (1949).
[9] E. T. Jaynes, *Ferroelectricity* (Princeton University Press, Princeton, 1953).
[10] See for example, J. C. Southard and F. G. Brickwedde, J. Am. Chem. Soc. **55**, 4378 (1933).

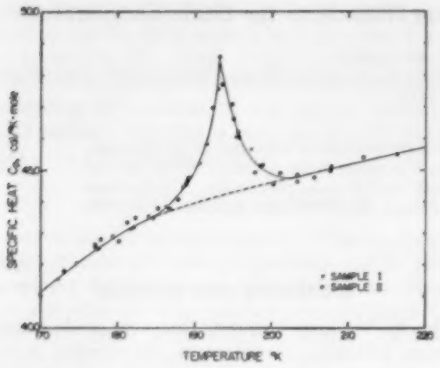


FIG. 1. Specific heat of $Cd_2Nb_2O_7$.

respectively, in good agreement with the values obtained by other investigators.[11–13]

PREPARATION OF MATERIAL

Two separate ceramic samples of $Cd_2Nb_2O_7$ were prepared by pre-firing milled mixtures of $CdCO_3$ and Nb_2O_5 at 1150°C, remilling and firing at 1250°C. The lattice constants, determined from powder photographs, were $a=10.371$ A for both sample I and sample II. These values are consistent with the values of a obtained from other specimens prepared in this laboratory.[5] Dielectric measurements revealed peak values of the dielectric constant at 191.4°K and 191.8°K for samples I and II, respectively.

It will be noted that the transition temperatures, T_c, of these samples are higher than any reported thus far. Hulm[3] has pointed out that the Curie temperature of $Cd_2Nb_2O_7$ depends on the temperature at which the specimen is fired. A higher firing temperature results in the lowering of the Curie point; unequal rates of evaporation of $CdCO_3$ presumably lead to slight variations in the stoichiometric formula.

RESULTS

The experimental specific heat data for $Cd_2Nb_2O_7$ between 170°K and 220°K were obtained from measurements performed on two calorimeter loadings (66.2 g of sample I and 69.7 g of sample II). Heating periods of 4, 5, 8, and 10 minutes were used, employing a heating rate of 0.36 watt. This heating rate produced an average temperature rise of 0.44°K per minute. The values of C_p are plotted as a function of temperature in Fig. 1. The values obtained from sample I are indicated by crosses (×), and those obtained from sample II by circles (○). The points lie on a smooth curve to within 0.5%, and the agreement of the points determined from the two samples is within the accuracy of the apparatus. The transition temperature, determined by the maximum value of C_p, is at 193.2°K, in good agreement with T_c determined from the dielectric measurements. Assuming that the normal specific heat can be taken as the broken curve shown in the figure, the value of the transition energy is 18±2 cal/mole, and the associated entropy change is 0.09±0.01 cal/°K-mole.

TABLE I. A comparison of the calculated and experimental values of the spontaneous polarization for Rochelle salt, KH_2PO_4, $BaTiO_3$, and $Cd_2Nb_2O_7$.

Material	ΔS, $\frac{\text{cal}}{\text{°K-mole}}$	C, °K	ρ/M, cc^{-1}	P_{calc}, $\frac{\mu\text{coul}}{\text{cm}^2}$	P_{exp}, $\frac{\mu\text{coul}}{\text{cm}^2}$
Rochelle salt	<0.007	180	1.8/282	<0.27	0.25
KH_2PO_4	0.72	260	2.3/136	5.4	5
$BaTiO_3$	0.12	12 700	6.1/233	24.0	20
$Cd_2Nb_2O_7$	0.09	5600	6.2/523	7.5	10

DISCUSSION

The transition entropy change, the maximum value of the spontaneous polarization P attained at low temperatures, and the Curie constant C (defined below) have been related theoretically by the equation $P^2=2C\Delta S$. This relation was derived by Jaynes[9] from the postulates of the Devonshire theory for $BaTiO_3$.[8] The derivation depends on the particular form of the free energy, which for zero stress may be expanded in even powers of the polarization. Moreover, it is assumed (a) that this free energy function is valid above and below the Curie temperature; (b) that the coefficients of the terms of order P^4 and higher are temperature independent and small compared with the coefficient of P^2; and (c) that the reciprocal susceptibility χ^{-1}, which is the coefficient of P^2, is a linear function of temperature: i.e., $\chi^{-1}=(T-T_c)/C$, where C is the Curie constant. Although general validity of this equation has not been established, it describes quite well the relation between ΔS, P, and C for Rochelle salt, KH_2PO_4, and $BaTiO_3$.

The equation may be rewritten in the form $P^2=9.3\rho C\Delta S/M$, where ρ is the density in g/cc and M is the molecular weight in g. This gives P in μcoul/cm^2 when ΔS is given in cal/°K-mole. For $Cd_2Nb_2O_7$, $M=523$ g, and $\rho=6.2$ g/cc as calculated from the lattice constant. Taking $\Delta S=0.09$ cal/°K-mole and $C=5600$°K, which is the average of all the values of C reported thus far, we obtain $P\sim7.5$ μcoul/cm^2. The experimental value of the spontaneous polarization has been reported as $P=10$ μcoul/cm^2.[5] For the sake of comparison, the experimental and calculated values of P for Rochelle salt, KH_2PO_4, $BaTiO_3$, and $Cd_2Nb_2O_7$ are listed in Table I.

The agreement between the experimental value of P and the value calculated from the experimental values

[11] G. Shirane and A. Takeda, J. Physiol. Soc. Japan 7, 1 (1952).
[12] J. Volger, Phillips Research Repts. 7, 21 (1952).
[13] S. S. Todd and R. E. Lorenson, J. Am. Chem. Soc. 74, 2043 (1952).

of ΔS and C has twofold significance. First, it indicates the essential correctness of all three of the experimental values. Second, it provides an additional example, in what may well become a new subclass of ferroelectrics, of the applicability of the equation $P^2=2C\Delta S$. In the light of the general agreement of this equation in previously studied ferroelectrics, it appears that a single free-energy function theory may be capable of explaining ferroelectricity in general.

The authors wish to thank Dr. Gen Shirane and Dr. Franco Jona for their helpful discussions, and Mr. John McLaughlin for preparation of the specimens and assistance in making the specific heat measurements.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Influence of Alternating Electric Fields on the Light Emission of Some Phosphors

I. T. Steinberger, W. Low, and E. Alexander
Department of Physics, The Hebrew University, Jerusalem, Israel
(Received April 4, 1955)

The influence of an alternating electric field on the fluorescence and phosphorescence of various ZnS and ZnS:CdS phosphors was investigated. The light pulses and the average brightness of the luminescence were measured simultaneously. The wave pattern during fluorescence consists of two different pulses per cycle. The two pulses differ in their respective amplitudes, time dependence, and shape. The pulses corresponding to positive polarity of the illuminated electrode show a similar behavior for all phosphors. The second pulse differs for the various phosphors. Voltage dependence of the amplitudes and light sums per cycle were found to be different for the two pulses. These two pulses must be due to two different processes and suggest surface effects.

INTRODUCTION

THE effects of alternating electric field on the luminescence of powdered nonelectroluminescent phosphors have been investigated by a number of authors. As early as 1920, Gudden and Pohl[1] found that on application of electric fields some of their photoconducting phosphors emitted momentary light flashes. Destriau and his co-workers[2–5] have made extensive investigations of the various influences of the electric field on the luminescence of some types of ZnS phosphors. They have studied in particular the changes in the average brightness of luminescence as a function of the voltage and temperature both during excitation and phosphorescence.

Olson[6] has investigated the quenching of the phosphorescence caused by the ac field as a function of the frequency for one particular phosphor (ZnS:CdS:Cu). Matossi and Nudelman[7] have studied the quenching during excitation as a function of the frequency for another particular phosphor (ZnS:Cu). Matossi[8] has attempted to explain the effects observed during excitation by extending the theory of Randall and Wilkins[9] on luminescence in phosphors. By including assumptions that the electric field causes emptying of traps as well as draining off of electrons by field-induced nonradiative transitions, he seems to account for the momentary illumination, subsequent quenching, and superposition of a ripple of frequency twice that of the applied voltage. It will be shown in this paper that there are a number of important effects which make it doubtful whether this theory in its present form can be applied to the phosphors we have investigated.

There has recently been a considerable amount of experimental and theoretical work on electroluminescent phosphors.[10] It should be emphasized, however, that explanations which may be valid in cases of electroluminescence do not necessarily account for the various effects of electric fields during fluorescence and phosphorescence. The hope of practical applications of electroluminescence has induced many workers to concentrate mainly on this phase of research. An understanding of the influence of electric fields on the usual nonelectroluminescent phosphor is more general in nature and may shed light on electroluminescence as well.

This series of experiments was undertaken in the hope of finding unifying features in the wealth of experimental data on the influence of ac fields on powdered phosphors. This paper is concerned with phosphors of the type ZnS and ZnS:CdS with Cu and Ag activators. In a later publication, we intend to report results on the effects of ac fields on doubly activated infrared-stimulated phosphors.

[1] G. Gudden and R. Pohl, Z. Physik **2**, 192 (1920).
[2] G. Destriau, J. Chim. Phys. **34**, 117, 462 (1937).
[3] G. Destriau, Phil. Mag. **38**, 700, 784, 880 (1947).
[4] G. Destriau and J. Mattler, J. Phys. radium **11**, 529 (1950); **13**, 205 (1952).
[5] G. Destriau, J. Appl. Phys. **25**, 66 (1954).
[6] K. W. Olson, Phys. Rev. **92**, 1323 (1953).
[7] F. Matossi and S. Nudelman, Phys. Rev. **89**, 660 (1953).
[8] F. Matossi, Phys. Rev. **94**, 1151 (1954).
[9] R. T. Randall and M. H. F. Wilkins, Proc. Roy. Soc. (London) **A184**, 390 (1945).
[10] A discussion of work in electroluminescence will be found in Brit. J. Appl. Phys., Supplement No. 4 (1954).

We have studied the various effects of a 50-cps ac field during both excitation and phosphorescence on the same phosphor. These effects were investigated as a function of the phosphor composition. Particular attention was paid to the shape of the light pulse, the light sum per pulse, and the amplitude, as functions of the time after the beginning of the application of the field and of the applied potential difference.

As will be shown later, our results indicate that there are two light pulses per cycle during both excitation and phosphorescence, each with different characteristics. During excitation these two pulses have different decay times, different light sums, and sometimes different shapes. All these effects are a function of the applied voltage, the function being different for the two pulses. It is probable, therefore, that there must be at least two mechanisms responsible for the emission of these two pulses. No simple model so far proposed can explain all the observed phenomena.

EXPERIMENTAL METHOD

The purpose of the experiments was to measure the slow changes in the average light output as well as the periodic light ripple. The light output was, therefore, measured by two methods:

(1) The light pulses from the phosphor cell were measured by a 1P21 photomultiplier tube, operated from a stabilized power supply. Suitable absorbing filters could be placed in front of the photomultiplier tube. The output was brought to the Y-axis of a cathode-ray oscilloscope equipped with a dc amplifier. The time base was synchronized to the line voltage. The pattern on the scope was recorded by means of an oscillo-record camera. The film of the camera could be moved with various constant speeds in the Y-direction of the scope.

(2) The above method was particularly suitable for measuring the characteristics of the light pulses but was not convenient for recording the slow changes of average brightness. For the latter purpose the output of the photomultiplier was fed to a moving coil galvanometer. The deflections were registered on photographic paper attached to a drum which was rotated at constant speed by means of a timing motor.

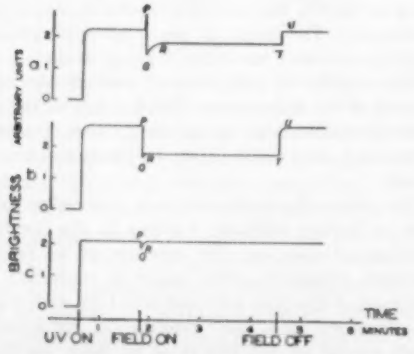


FIG. 1. Effect of the alternating electric field on the average brightness during excitation for the phosphors (a) ZnS:Cu (RCA F-2040), (b) ZnS:Ag (RCA F-2030), (c) 20ZnS:80CdS:Ag (RCA F-2038).

The phosphor cells were similar in construction to those reported by other observers.[3,10–12] The phosphors were embedded in thermal-setting polystyrene or in araldite. Conducting glass was prepared by the method of Gomer,[13] using microscope cover glasses in order to minimize ultraviolet absorption. Contact to the conducting glass was made by means of a thin aluminium foil and aluminium strips fastened to the foil at the edge of the glass.

Cells were constructed in which both electrodes were made of conducting glass, and others in which one electrode was of brass. In some cells the phosphor matrix made direct contact with the electrodes, in others thin mica sheets were placed between the matrix and the electrodes.

The phosphor was excited by ultraviolet light (mainly 3650 A) from a stabilized mercury vapor lamp operated from a dc supply. A camera shutter and suitable filters could be placed in front of this lamp. The exciting light was focused on the conducting glass. The emitted light, passing through this glass, fell on the photomultiplier cathode. Measurements were made only when the phosphor was fully excited (in general after two minutes of excitation).

A condenser of capacity 0.1 μf was connected across the phosphor cell in order to prevent possible voltage transients caused by switching the field on and off.

Phase measurements were made by means of the Z amplifier of the scope. A pulse-sharpening circuit was connected between the ac source and the Z amplifier. After making corrections for the phase difference caused by the pulse sharpener and amplifier, one could measure the phase difference between the maxima of the applied voltage and the maxima of the light pulses. Sufficient time was allowed between experiments for the decay of the phosphor.

EXPERIMENTAL RESULTS

The phosphors investigated were of the type (a) ZnS:Cu (RCA F-2040), (b) 80ZnS:20CdS:Cu (RCA F-2042), (c) ZnS:Ag (RCA F-2030), (d) 80ZnS:20CdS:Ag (RCA F-2032), (e) 20ZnS:80CdS:Ag (RCA F-2038). The various effects observed have been tested in dielectric cells made of polystyrene and araldite. The results are reproducible and those reported are an average of many cells. The following general conclusions can be inferred from the experimental results.

1. On application of an alternating electric field during *excitation*, the average brightness is increased

[11] S. Roberts, J. Opt. Soc. Am. **42**, 850 (1952).
[12] Payne, Mager, and Jerome, Illum. Eng. **45**, 688 (1950).
[13] R. Gomer, Rev. Sci. Instr. **24**, 992 (1953).

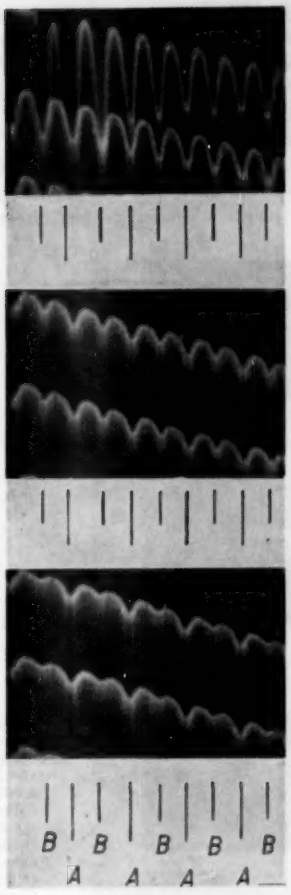


FIG. 2. Motion pictures of light pulses during excitation of ZnS:Cu for various times after application of the field. ($V=630$ volts.) (a) $n>0$, (b) $n>80$, (c) $n>1000$. (n in units of 1/50 sec.) The maxima of the pulses are directed downwards. Pulses A correspond to positive polarity of illuminated electrode.

initially for all phosphors investigated. This sharp increase [at P in Fig. 1(a)] lasts only a very short time, and is followed by a momentary dip [at Q in Fig. 1(a), (b), (c)]. In most phosphors, this increase lasts less than half a second and the galvanometer record does not show it [Fig. 1(b) or 1(c)]. Photographic records, however, indicate this rise. After this transitory period (time $P-Q$) the average brightness for most phosphors reaches a new equilibrium level ($R-T$), lower than the luminescent level before application of the field. The phosphor 20ZnS:80CdS:Ag is an exception and shows only a momentary dip and no stationary quenching [Fig. 1(c)].

An ac wave is superimposed on the average brightness level. The shape and amplitude of these light pulses change rapidly in the transitory period and have constant characteristics in the stationary state (time $R-T$). Figure 2 shows a typical photograph of the light pulses for ZnS:Cu phosphor during the transitory period [Fig. 2(a)] and stationary period [Fig. 2(c)].

2. The various phosphors show no clear-cut relationship between the height of the pulses and the amount of quenching in the stationary period. There is some indication that in phosphors which show large quenching the corresponding pulses are small, and conversely in those which show small quenching the amplitudes are large. For example, ZnS:Ag shows large quenching and very small pulses, while 20ZnS:80CdS:Ag shows no quenching (except momentarily) and larger ac pulses. But ZnS:Cu shows significant quenching and very large pulses.

3. In the transitory period, all phosphors show two pulses per cycle. In the stationary period ($R-T$ in Fig. 1) there are either one or two pulses per cycle, depending on the applied voltage. At very high voltages two pulses per cycle always appear. The behavior of the two pulses depends on the polarity of the illuminated electrode and not on the direction of observation. We distinguish between pulse A observed when the illuminated electrode is positive and pulse B when the illuminated electrode is negative.

4. The two observed pulses per cycle differ in the stationary state considerably in their characteristics: (a) The amplitude A and light sum per cycle S_A are not equal to B and S_B.[14] This is seen in Fig. 2 for the phosphor ZnS:Cu. (b) Their dependence on applied voltage is different (see Figs. 3 and 4). (c) The amplitude B always starts to appear at a higher threshold voltage than A. In some phosphors (as in 80ZnS:20CdS:Cu) the pulse B is missing for all voltages except extremely high voltages close to dielectric breakdown. (d) For low, medium, and in some cases even for high voltages, the A component is larger than the B component.

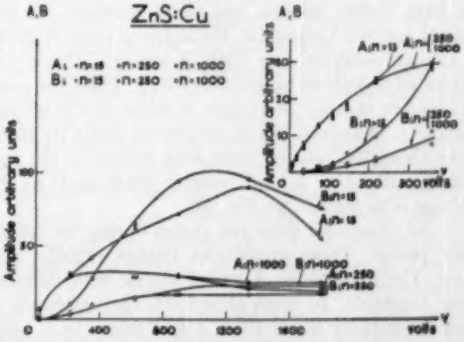


FIG. 3. Amplitudes of light pulses A and B during excitation as a function of the potential difference for ZnS:Cu (RCA F-2040). The insert is an enlarged portion of the same graphs for low potential differences.

[14] The amplitude is measured by drawing a straight line connecting adjacent minima and measuring the vertical distance from the apex to this line. The light sum is similarly measured by integrating the area bounded by this line and the outline of the pulse.

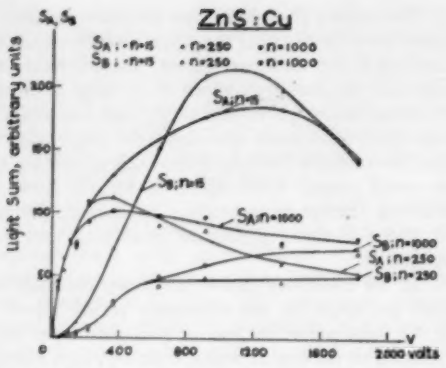


FIG. 4. Light sum per cycle, S_A and S_B during excitation as a function of the potential difference for ZnS:Cu (RCA *F*-2040).

5. In the stationary state the amplitude A and the light sum per cycle S_A show a similar dependence on the voltage for all phosphors. Figures 3 and 4 show the typical variations with voltage in the case of ZnS:Cu. The pulse A has a relatively low threshold voltage (differing for the various phosphors in magnitude), and rises with the voltage to a maximum and subsequently decreases. The amplitude B has a higher threshold voltage and rises slowly in an S-shaped curve to a saturation value. At very high voltages, the two components A and B are nearly equal (ZnS:Cu), and in some phosphors B may even be larger than A. At extremely high voltages, both A and B decrease with further increase of voltage.

6. The shape of the pulse A is a function of the voltage and of the time. With increase of voltage the pulse becomes more and more peaked and narrower at its base (broad minima and sharp maxima). In the copper activated phosphor, the shape of pulse B shows a similar behavior (see Figs. 2 and 5). The silver-activated phosphors behave differently. There the pulse B becomes broader and larger in shape as the voltage increases. Moreover, the B component shifts its phase from about 180° out of phase with respect to the preceding pulse A to a much smaller phase angle as the voltage is increased (see Fig. 6).

7. All phosphors show two pulses during the transitory period. These amplitudes change rapidly with time. The amplitudes A and B decay with different time constants. In some phosphors, the amplitude B may be initially larger than A and decay to a value smaller than A. Figure 7 shows the amplitudes A and B as a function of time (in units of 1/50 sec) both in the transitory and stationary state for ZnS:Cu.

8. On application of alternating electric fields during *phosphorescence*, the average brightness increases initially and is followed by some quenching. Two light pulses A and B are superimposed and are always present for all phosphors. The amplitudes A and B decay with nearly equal time constants which are much smaller than that of the normal phosphorescence. For some phosphors A is larger than B; for others (e.g., ZnS:Cu) B is larger [(see Fig. 7(b)]. There is a reversal in the relative magnitudes of A/B with time in some phosphors.

9. The results are qualitatively similar for cells made of various dielectrics and are independent of whether the phosphor matrix makes direct contact with the

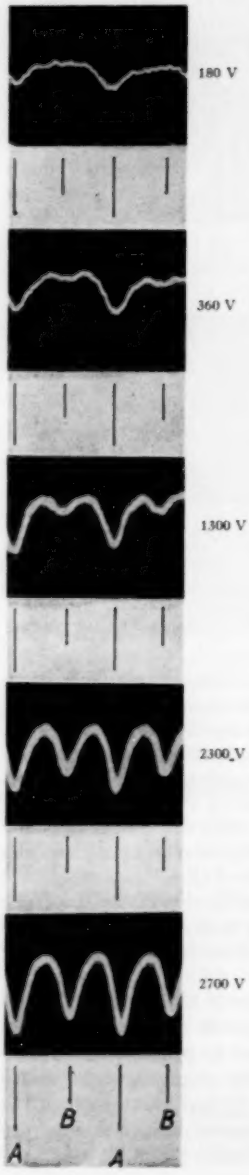


FIG. 5. Light pulses A and B during excitation as a function of the potential difference, in the steady state region. Note the changes in shape of the pulses A and B. ZnS:Cu (RCA *F*-2040). Maxima directed downwards.

electrodes or is separated by means of mica sheets from these electrodes. The effects are independent of whether the conducting glass or the metal electrode is grounded.[15]

10. If the electric field is applied before the excitation, it takes a few seconds until the amplitudes A and B are built up to their respective magnitudes.

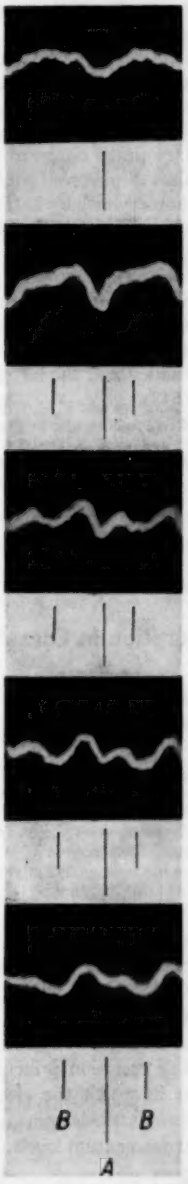


FIG. 6. Light pulses (A and B) during excitation as a function of the potential difference, in the steady state region, for 80ZnS:20CdS:Ag (RCA F-2032). Maxima directed downwards.

[15] See A. Luyckx and A. S. Stockking, reference 10, paper 10.

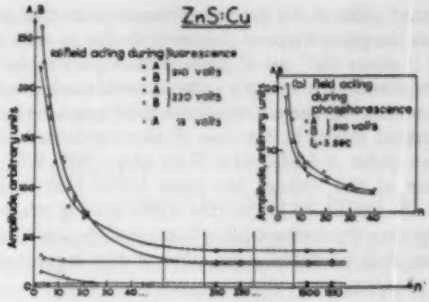


FIG. 7. Amplitudes of the light pulses A and B of ZnS:Cu (RCA F-2040) as a function of time when the field is acting (a) during excitation; (b) during phosphorescence. The field was applied at $t_0=5$ sec after cessation of the excitation.

DISCUSSION

Our results indicate that the two observed amplitudes A and B are due to two different effects, because the dependence of these pulses on voltage and on time is very different for all phosphors. The observation of the two effects is possible, because the exciting radiation is strongly absorbed in the phosphor. This was tested, for example, on a symmetrical phosphor cell (both sides having conducting glass electrodes) containing 80ZnS:20CdS:Cu, and it was found that less than 1% of the exciting radiation (3600 A) was transmitted. On the other hand, the transmission of light of wavelength 6000 A, corresponding to that of the emitted light of this phosphor, was about 25%. This indicates that in our experiments the polarity of the amplitudes A and B is to be defined by the polarity of the ultraviolet-illuminated electrode. The amplitude A corresponds to a positive polarity of this electrode. This was checked by changing the direction of the exciting illumination, the direction of the observation remaining the same. It was found that, in this case as well, A occurred when the illuminated electrode was positive. We can, therefore, conclude that the effects we observed occurred in the main near the illuminated electrode.

The amplitudes A and B differ not only in magnitude but also in character. The difference between the voltage thresholds of the two amplitudes could possibly be explained by absorption effects. However, the very different voltage dependence, and the fact that in some phosphors the ratio of the two amplitudes A/B changes from >1 to <1 with increase of voltage, precludes such an explanation. In a similar way, one cannot explain the differences in time dependence and phase and shape changes without the assumption of at least two different processes.

Pulse A shows very similar characteristics for all phosphors in its voltage dependence, differing only by scale factors from one phosphor to the next. Pulse B, however, has different characteristics for copper- and silver-activated phosphors, either case differing from

those of pulse A. In the case of copper-activated phosphors the pulse shape of B is very similar to that of A and is about 180° out of phase with respect to A. The voltage and light sum per pulse curve is reminiscent of the voltage to current relationship of a space-charge-saturated diode. In the case of silver-activated phosphors, pulse A leads pulse B by about 90°. With increase of the voltage the pulse width gets broader and the amplitude larger (the width getting relatively wider than the corresponding increase of the amplitude). Pulse B is probably a function of the impurities in the phosphor.

There are similarities between the effects during the transitory period and those during phosphorescence. In both cases one observes rapid changes in the average brightness and always two amplitudes per cycle. For short persistent phosphors the transient effects during both phosphorescence and excitation are of short duration. Apparently this effect can be connected with emptying of traps. The subsequent quenching may also be related to changes in transition probabilities (or radiationless transitions) due to the influence of the electric field.

From the theory of Matossi[8] it follows that two pulses of equal amplitude per cycle should be observed. We found at some voltages only one pulse. Furthermore, as explained before, the different behavior of A and B proves that whenever there are two pulses per cycle, they are due to two different effects.

We think that the appearance of the two amplitudes suggests that surface effects are involved.

Experiments in electroluminescence have shown that there exist in general two pulses of light per cycle which are out of phase with voltage.[16–18] At certain voltages, two secondary light pulses are found about 90° out of phase with the main peaks.[16,19] For silver-activated phosphors, we found that the pulse B was about 90° out of phase with the pulse A, whereas for copper-activated phosphors B was about 180° out of phase. Zalm *et al.*[19] found no secondary light peaks for copper-activated phosphors. It is tempting to correlate the secondary peaks observed in electroluminescence with the pulses B observed during excitation. In this case, the nonobservance by Zalm *et al.* of the secondary peak in copper-activated phosphors in electroluminescence could possibly be explained by a masking of the secondary peak by the next main peak (because of about 180° phase shift). Apparently no such correlation can exist, however, between the main peaks in electroluminescence and the pulses A, as the former are emitted near the cathode[16] and the latter near the anode.

[16] W. W. Piper and F. E. Williams, Phys. Rev. **87**, 191 (1952).
[17] J. F. Waymouth and F. Bitter, Phys. Rev. **95**, 941 (1954).
[18] See also papers in reference 10.
[19] Zalm, Diemer, and Klasens, Philips Research Repts. **9**, 81 (1954).

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Paramagnetic Resonance Absorption in Glass*†

R. H. SANDS‡
Department of Physics, Washington University, St. Louis, Missouri
(Received March 2, 1955)

Paramagnetic resonance absorptions occurring with spectroscopic splitting factors of 4 and 6 were noted in all glass samples studied. These were apparently anisotropic g-values for a specific impurity and the peaks recorded are interpreted as being the result of a distribution of Stark fields over all directions in the glass. Resonances were also recorded having g-values near 2. Nearly all of these showed hyperfine structures characteristic of the known paramagnetic additions in the glasses. The absorption line observed for cupric ions in lattice modifying positions in a soda-lime-silica base glass has been quantitatively explained and the parameters compared favorably with those resulting for copper in the Tutton salts.

I. INTRODUCTION

CHANCE observations of paramagnetic resonance absorption in glass have led to a detailed study of such absorption in numerous glass samples of varying composition furnished by the Corning Glass Works, Corning, New York. The results have indicated possible applications of this technique to glass technology as well as in the study of the vitreous state.

It has long been known that if an atom has a partially filled electronic shell, there will exist a permanent electronic magnetic dipole moment. Hund's rules[1] and the Pauli exclusion principle allow one to predict the manner in which the electrons in the unfilled shell, with their intrinsic spins, will fill the available orbital angular momentum levels.

* Supported in part by the U. S. Office of Scientific Research, Air Research and Development Command.
† Submitted to the graduate board of Washington University, St. Louis, Missouri, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
‡ Present address: Physics Department, Stanford University, Stanford, California.

[1] F. Hund, *Linienspektren und periodisches System der Elemente* (Verlag Julius Springer, Berlin, 1927), p. 124.

When the atom or ion is placed in a lattice position in a solid, the surrounding diamagnetic neighbors produce an electrostatic potential at the ion in question. This potential causes a splitting of the orbital levels in accord with group theoretical considerations.[2] The magnitude of this Stark splitting varies with environment but is normally several thousand wave numbers for iron-group ions. Since this is much larger than the thermal energy (200 cm^{-1} at 300°K), one need consider only the ground state to first approximation. Should this ground state be a nondegenerate orbital state, then it must have the total spin degeneracy of the system (the splitting of the spin degeneracy is small). If the zero-order Stark interaction has such symmetry as to leave the ground orbital state degenerate, then this ground state will split according to the theorem of Jahn and Teller[3] leaving only the Kramers degeneracy. It is reasonable to conclude that the ground state will be a doublet if the number of electrons is odd and a singlet if the number is even (the next higher state usually being some 100 cm^{-1} away). For spin doublet ground states it has been shown by Pryce that the spin-orbit coupling between states and the electronic Zeeman interaction with an applied magnetic field H can be considered to give two ground-state energy levels:

$$E_{\pm}=\pm\tfrac{1}{2}g\mu_B H, \tag{1}$$

where $g=(g_{\parallel}^2\cos^2\theta+g_{\perp}^2\sin^2\theta)^{\frac{1}{2}}$, μ_B is the Bohr magneton, and θ is the angle between the applied magnetic field and the axis of symmetry of the crystalline electric field. For a $^2D_{\frac{5}{2}}$ state under cubic plus tetragonal symmetry, Polder shows[4] that

$$g_{\parallel}=2(1-4\lambda/\Delta),\quad g_{\perp}=2(1-\lambda/\Delta'), \tag{2}$$

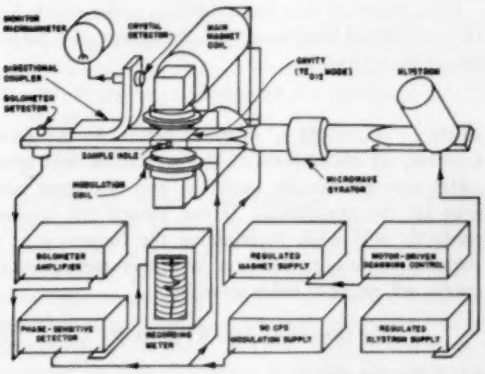


FIG. 1. Block diagram of the microwave absorption spectrometer.

where λ is the spin-orbit coupling constant and Δ and Δ' are the energy differences between the ground state and two higher Stark levels.

The interactions involving the nuclear moment can be written, neglecting the nuclear quadrupole moment, as

$$\mathcal{H}_n=g_n\mu_n\mathbf{I}\cdot\mathbf{H}+a\mathbf{L}\cdot\mathbf{I}+b\{3r^{-2}(\mathbf{I}\cdot\mathbf{r})(\mathbf{S}\cdot\mathbf{r})-\mathbf{I}\cdot\mathbf{S}\}+c\mathbf{I}\cdot\mathbf{S} \tag{3}$$

where c includes the probability that the electron is at the nucleus.[5] Such an interaction when added to the previous interactions yields the energy levels:

$$E_+(M_I)=\tfrac{1}{2}g\mu_B H+\left[\left(\frac{A_{\parallel}g_{\parallel}}{2g}-g_n\mu_n H\right)^2\cos^2\theta+\left(\frac{A_{\perp}g_{\perp}}{2g}-g_n\mu_n H\right)^2\sin^2\theta\right]^{\frac{1}{2}}M_I,$$
$$E_-(M_I)=-\tfrac{1}{2}g\mu_B H-\left[\left(\frac{A_{\parallel}g_{\parallel}}{2g}+g_n\mu_n H\right)^2\cos^2\theta+\left(\frac{A_{\perp}g_{\perp}}{2g}+g_n\mu_n H\right)^2\sin^2\theta\right]^{\frac{1}{2}}M_I. \tag{4}$$

For a $^2D_{\frac{5}{2}}$ state, if one uses Eq. (2),

$$A_{\parallel}=(g_{\parallel}-2)\langle a\rangle+(3/7)(g_{\perp}-2)\langle b\rangle-(4/7)\langle b\rangle+\langle c\rangle, \tag{5}$$
$$A_{\perp}=(g_{\perp}-2)\langle a\rangle-(3/14)(g_{\perp}-2)\langle b\rangle+(2/7)\langle b\rangle+\langle c\rangle.$$

The above five equations are sufficient to describe the resonances reported in detail in this paper. The general shapes of the resonance absorption curves can be obtained by considering the amorphous character of the samples studied, as will be shown.

II. VITREOUS STATE

The physics of glass is complicated by the fact that this vitreous state is not in thermal equilibrium. Rather such a state exists only by virtue of negligible rates of crystallization.

The glasses of interest here are those based on silica. Furthermore, all glass samples employed have been the product of fusing together various oxides. This does not imply that these oxides exist as such in the vitreous state, however. They fulfill certain functional purposes in the glass, two of which are labeled "network-formers" and "network-modifiers," by which is meant that some oxides have a coordination number for the metallic ion which is suitable for forming long oxide "chains," while other oxides present coordination numbers not suitable for such a formation and hence must either allow their metallic ions to fill interstitial positions in the network or join with conjugate oxides

[2] H. Bethe, Ann. Physik **3**, 133 (1929).
[3] H. A. Jahn and E. Teller, Proc. Roy. Soc. (London) **A161**, 220 (1937).
[4] D. Polder, Physica **9**, 709 (1942).
[5] For a comprehensive discussion of the appropriate Hamiltonians see B. Bleaney and K. W. H. Stevens, Repts. Progr. Phys. **16**, 108 (1953).

to form a binary network. In a silica glass these functions demand that a metallic ion either lie inside a tetrahedron of oxygen atoms or rest in an interstitial position where its coordination number may be different from two.

The position which an ion assumes in the glass is of considerable importance since the optical absorption of the ion depends upon its surroundings as noted in Sec. I. In fact, it is just this dependence that is of interest here.

III. EXPERIMENTAL ARRANGEMENT

The apparatus employed in studying the effect of the environment on the ground state spin levels of the paramagnetic ions in the various glasses is outlined in Fig. 1. This paramagnetic resonance (PMR) spectrometer was designed to sweep the applied magnetic field from zero to 5000 gauss while the paramagnetic sample is bathed continuously in electromagnetic radiation at 9000 Mc/sec. The power absorbed by the sample from the radiation bath is detected and recorded *versus* the applied magnetic field.

The circuit consists of a reflex klystron oscillator feeding microwave power through a ferromagnetic isolator (gyrator) into a transmission cavity (operating in the TE_{012} mode) which contains the sample under investigation. The power transmitted by the cavity is carried to a barreter where any change of power results in a change of dc resistance of the barreter. By virtue of a constant direct current furnished to the barreter, this change of resistance manifests itself as a change in voltage across the barreter which is then amplified through an audio-amplifier, fed into a phase-sensitive detector and finally onto an Esterline-Angus recorder. The applied magnetic field is modulated at a ninety-cycle rate and a phase-coherent reference signal is applied to the grids of the phase-sensitive detector switching tube so as to allow only the resonance signal to pass through the time constant across the switching tube.

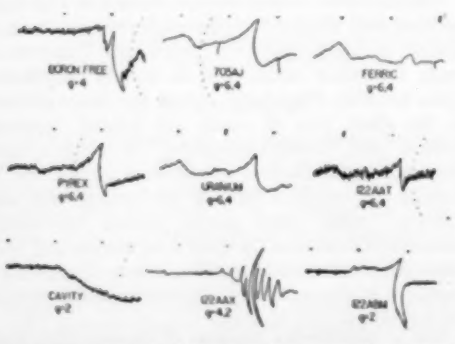


FIG. 2. Representative derivatives of paramagnetic resonance absorption lines in glasses.

The circuit for the magnetic field sweep consists of a constant current supply whose reference voltage may be swept continuously by means of a helipot. The clock motor in the recorder drives both the chart and the reference voltage helipot. The resulting current sweep through the magnet coils is quite linear with time; over the ranges employed in the following experiments, the field sweep is linear as well.

IV. EXPERIMENTAL RESULTS

Figure 2 shows a representative group of derivatives of the paramagnetic resonance absorption curves recorded for the glass samples listed in Table I. In nearly all cases, save where dielectric absorption ruled otherwise, the samples were rectangular sections of 0.5-cm² cross-sectional area, extending across the total width of the microwave cavity at a region where the electric field was a minimum. Table I displays the values of the spectroscopic splitting factors (g-values) observed for each sample surveyed. It is of interest to note that in all cases there existed resonances having apparent anisotropic g-values of 4 and 6. That these are anisotropic g-values for a particular paramagnetic impurity in the glasses is, of course, not immediately obvious. However, the shapes of the integrated curves are in good accord with this assumption, as will be shown.

Prior to any computations, one may make a few qualitative remarks. Experiment shows that the spectra observed do not seem to alter upon working the glass. This, perhaps, indicates that impurities are responsible and that nothing analogous to F-centers is operative. Such a conclusion is further substantiated by the fact that for glasses made from the same base, the line intensities remain the same; but for varying bases one finds varying intensities.

Let us suppose that the conditions under which Eq. (1) was derived hold true. Then the energy of a $\Delta m=1$ transition is given by

$$h\nu = g\mu_B H, \tag{6}$$

where $g=(g_{\parallel}^2\cos^2\theta+g_{\perp}^2\sin^2\theta)^{\frac{1}{2}}$ as before. In the glass, however, all orientations of the crystalline field symmetry axis are equally probable, and one must sum over all the absorptions. If one sweeps the applied magnetic field while maintaining the frequency at a constant value ν_0 as was done here, then absorption of energy will occur at fields given by

$$H=(h\nu_0/\mu_B)(g_{\parallel}^2\cos^2\theta+g_{\perp}^2\sin^2\theta)^{-\frac{1}{2}}, \tag{7}$$

for each value of θ.

The number of spins having for their environment an electric field symmetry axis at an orientation with respect to the applied magnetic field between θ and $\theta+d\theta$ is given by

$$dN=(N_0/2)\sin\theta d\theta, \tag{8}$$

TABLE I. Apparent g-values observed in glasses. The numbers in parentheses indicate the percentage by weight of the individual paramagnetic impurities added to the purest soda-lime-silica base glass available. Resonances marked "Str" show hyperfine structures characteristic of the known paramagnetic additions.

Pyrex	6.2; 4.2	122ABA CuO, TaO, WO_3 (0.4%)	6.0; 4.2; 2(Str)
Nonex	6.4; 4.3	122ABB ZrO, Cb_2O_5 (0.4%) Nd_2O_3, Pr_2O_3 (0.2%)	6.0; 4.2
705AJ	6.0; 4.0	122ABC SnO, YtO, Ce_2O_3	6.0; 4.2
Fused quartz	6.2; 4.2	122ABD TiO_2, V_2O_3 (0.01%)	6.0; 4.2; 2(Str)
Neutron irrad. crystalline quartz	5.0; 4.4	122ABE Mn_2O_3, Cr_2O_3 (0.01%)	6.0; 4.2; 2(Str)
Boron free	6.0; 4.3	122ABF	6.0; 4.2
Lime	6.0; 4.3	122ABG NiO, MoO_3, (0.01%)	6.0; 4.2
Soft glass No. 1	6.0; 4.3	122ABH CuO, TaO, WO_3 (0.01%)	6.0; 4.2; 2(Str)
Cobalt	7.0; 4.4	122ABI ZrO (0.01%) Nd_2O_3, Pr_2O_3 (0.005%)	6.0; 4.2
Uranium	6.0; 4.2	122ABJ SnO, Ce_2O_3, YtO (0.01%)	6.0; 4.2
122AAT ordinary base materials	6.0; 4.2	122ABK Fe_2O_3 (0.01%)	6.0; 4.2
122AAU As_2O_3 plus sulfates	6.0; 4.2	122ABL	6.0; 4.2
122AAV As_2O_3	6.0; 4.2	122ABM CuO (0.4%)	6.0; 4.2; 2(Str)
122AAW sulfates	6.0; 4.2	122ABN WO_3 (0.4%)	6.0; 4.2
122AAX TiO_2, V_2O_3 (0.4%)	6.0; 4.2; 2(Str)	122ABO TaO (0.4%)	6.0; 4.2
122AAY Mn_2O_3, Cr_2O_3 (0.4%)	6.0; 4.2; 2(Str)	122ABP pure lime, ordinary sand	6.0; 4.2
122AAZ NiO, MoO_3, (0.4%)	6.0; 4.2; 2(Str)	122ABQ pure sand, ordinary lime	6.0; 4.2

which becomes, from Eq. (7),

$$dN=(N_0/2)(4H_0^2/H^3)\times\{(g_{11}^2-g_\perp^2)[(2H_0/H)^2-g_\perp^2]\}^{-\frac{1}{2}}dH, \quad (9)$$

where $H_0=h\nu_0/2\mu_B$.

Figure 3 shows dN/dH from Eq. (9) plotted *versus* H together with an integrated curve obtained for 705AJ (the derivative curve from which this was taken is that shown in Fig. 2). If one allows the individual lines whose centers are distributed according to dN/dH to have a finite width, the agreement between theory and experiment can be made quite satisfactory.

The assumption that these lines are due to an as yet undetermined impurity having anisotropic g-values ($g_\perp\approx4$ and $g_{11}\approx6$) in its lattice site in glass appears to be consistent with experiment. It remains, perhaps, to demonstrate that such g-values can be expected to exist. The cobalt ion furnishes the best example. It has been observed to have values between 2.5 and 3.5 for $g_\perp$ and between 5.5 and 7 for g_{11}. In glass in relatively large concentration, it shows a $g_\perp$ of 4.4 and a g_{11} of 7 as noted in Table I. Further experimentation is needed, however, to identify the paramagnetic impurity observed here.§

V. CUPRIC ION RESONANCE IN GLASS

In addition to the resonances near $g=4$ and 6 which occurred in all glass samples, there were resonances located near $g=2$ which were characteristic of the known added paramagnetic impurities. An example of these resonances is that for Cu^{2+} in a lattice-modifying position in a soda-lime-silica base glass. The sample employed was labeled 122ABM, and its spectrum is recorded in Fig. 2. Figure 4 shows this derivative curve on rectangular coordinates together with its integrated absorption curve. One notes how much more sensitive the derivative is to structure than is its integrated counterpart. It shows two sets of four equally spaced peaks—one set centered about 2770 gauss and the other about 3130 gauss with considerably different spacings. From Sec. I, the conditions of a large Stark splitting of the orbital levels ($^2D_\frac{1}{2}$) together with a spin-orbit coupling and nuclear hyperfine interactions seem to fit the data.

Equations (4) and (5) were derived under these assumptions, neglecting the nuclear quadrupole moment. If one also neglects $g_n\mu_nH$ relative to $A_{11}/2$ and $A_\perp/2$, then

$$h\nu_{\frac{1}{2}\leftrightarrow-\frac{1}{2}}=g\mu_BH+2M_I(g_{11}^2\cos^2\theta+g_\perp^2\sin^2\theta)^{-\frac{1}{2}}\times(d^2\cos^2\theta+f^2\sin^2\theta)^{\frac{1}{2}}, \quad (10)$$

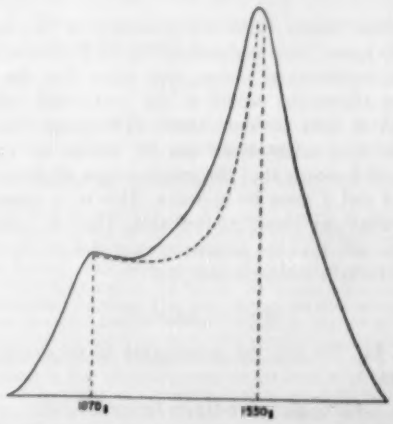


FIG. 3. Theoretical and experimental absorption curves for glass sample 705AJ. The dashed curve represents the plot of the theoretical distribution of line centers [dN/dH of Eq. (9)] *versus* H. The solid curve is the integrated absorption curve from a sample of 705AJ glass.

§ *Note added in proof.*—H. Weaver, Varian Associates, has informed the author that an increase in intensity of this anisotropic absorption has been observed in a soda lime glass containing 0.4% Fe_2O_3. This implies that Fe^{+++} may be responsible for these lines in all the samples studied but Fe^{++} is perhaps more likely.

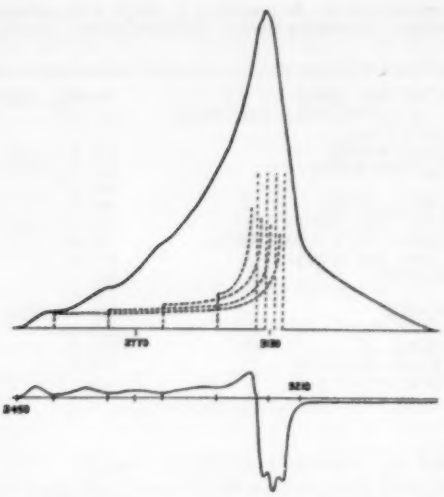


FIG. 4. Derivative and integrated absorption curves for Cu^{2+} in a lattice-modifying position in a soda-lime-silica base glass. The dashed curves are the theoretical line center distributions from Eqs. (8) and (11).

with $d=A_{\parallel}g_{\parallel}/2$ and $f=A_{\perp}g_{\perp}/2$. Defining $h\nu_0=2\mu_B H$,

$$H=2H_0/g-2M_I\mu_B^{-1}(g_{\parallel}^2\cos^2\theta+g_{\perp}^2\sin^2\theta)^{-1}\times(d^2\cos^2\theta+f^2\sin^2\theta)^{\frac{1}{2}}, \quad (11)$$

again, one expects that the axes of the Stark fields at the ion sites in the glass will be randomly oriented. As before, one may plot dN/dH *versus* H. The upper and lower limits in field are

$$H_U=2H_0/g_{\perp}-2Mf/g_{\perp}^2\mu_B, \quad (12)$$
$$H_L=2H_0/g_{\parallel}-2Md/g_{\parallel}^2\mu_B,$$

and these clearly fulfill the condition of two sets of equally spaced lines as shown in Fig. 4. If one now refers to the experimental curves, one notes that the most intense absorption occurs at the lower field value in the set of lines centered about 3130 gauss. Since all nuclear spin orientations can be treated as equally likely, this means that the relative signs of the parameters d and f must be opposite. This is in agreement with what one considers probable. That is, from Eq. (3) one sees that the parameters a and b are functions of the radial coordinate and that

$$a=b=g_n g\mu_n\mu_B/r^3. \quad (13)$$

From Eq. (5) and the assumption of no s-state promotion,

$$A_{\parallel}=[(g_{\parallel}-2)+3(g_{\perp}-2)/7-4/7]\langle a\rangle,$$
$$A_{\perp}=[(g_{\perp}-2)-3(g_{\perp}-2)/14+2/7]\langle a\rangle. \quad (14)$$

Taking the values $g_{\parallel}=2.32$ and $g_{\perp}=2.06$ from Fig. 4, one obtains $A_{\parallel}=-0.23\langle a\rangle$ and $A_{\perp}=0.33\langle a\rangle$. Thus, from Eq. (10) one might expect the parameters d and f to have opposite signs since the amount of promotion is certainly small.

From Fig. 4 and the preceding definitions one finds $g_{\parallel}=2.32$, $g_{\perp}=2.06$, $d=-548$ Mc/sec and $f=72.6$ Mc/sec. These data demand that $\langle c\rangle$ in Eq. (5) be negative. If one defines κ by the equation $\langle c\rangle=-\kappa\langle a\rangle$, then one obtains

$$A_{\parallel}=-(\kappa+0.23)\langle a\rangle,$$
$$A_{\perp}=(0.33-\kappa)\langle a\rangle. \quad (15)$$

From the values for d and f one finds $\kappa=0.26$. These experimental values are in excellent accord with the results reported for copper in the Tutton salts.[6,7]

CONCLUDING REMARKS

These resonances show clearly that there is definite short-range order in the lattice-modifying positions of many glasses, and the ability to investigate the symmetry of the individual lattice positions in the vitreous state should be of considerable aid to the glass technologist. With higher impurity content, resonances not reported for paramagnetic ions in Table I have become observable.[8]

The hyperfine structure observed for the titanium-vanadium loaded sample (122AAX) has been recorded for a vanadium loaded sample alone. Resonance studies on vanadium ions in solution indicate that the strong central absorption is due to vanadium ions in one lattice site while the eight-line spectrum is attributable to vanadium ions in another site. Whether these are also different ionization states has not been determined. The central absorption grows in the liquid state as the concentration of the vanadium salt is increased. Theoretical work is underway in an attempt to explain the width variation of the individual lines of the eight-line spectrum recorded in solution, and this should shed considerable light on the spectrum noted here in glass. The results of this latter work will be published shortly.

ACKNOWLEDGMENTS

This author wishes to thank Professor G. E. Pake for many helpful discussions and the Research Department of the Corning Glass Works for its cooperation and assistance in furnishing the loaded samples. Acknowledgment is also made to Dr. Jonathan Townsend for the design and construction of the apparatus employed.

[6] Bleaney, Penrose, and Plumpton, Proc. Roy. Soc. (London) **A198**, 406 (1949).
[7] A. Abragam and M. H. L. Pryce, Proc. Roy. Soc. (London) **A206**, 164 (1951).
[8] H. Weaver (private communication).

Recombination of Injected Carriers at Dislocation Edges in Semiconductors

J. P. McKelvey and R. L. Longini
Westinghouse Research Laboratories, Pittsburgh 35, Pennsylvania

(Received April 14, 1955)

The boundary conditions for carrier recombination at a lineage boundary in a semiconductor in the presence of a constant electric field are derived from kinetic considerations. A nonrecombination probability, or "transmission coefficient" T for a single charge carrier is assumed initially, and "recombination velocities" at the lineage boundary are expressed in terms of this quantity. A simple experiment to determine the transmission coefficient is suggested. The relationship between the transmission coefficient and the cross section for recombination is investigated, and a relationship derived which permits one to calculate the actual recombination cross section for a single dislocation, the density of dislocations in the lineage boundary having been experimentally determined.

I. INTRODUCTION

THE recent discovery of direct evidence of the existence of edge-type dislocations[1] at lineage boundaries in germanium[2] has aroused widespread interest in the properties of this type of crystal imperfection. It sometimes happens that two sections of germanium (or other) crystal are misoriented with respect to one another by a very small angle. At the boundary between two such sections exists a row of edge-type dislocations, the presence of which can be revealed, in the case of germanium, as a row of etch pits.[2]

It has been shown that recombination rates of injected carriers in germanium are increased by the presence of a high density of dislocations, and, in fact, that a single lineage line acts as a center of recombination for minority carriers in germanium.[3] The experiments of Vogel, Read, and Lovell are subject to some errors, however, due to the fact that their experimental arrangement only approximately represents the boundary value problem whose solution they utilize, and also because the boundary conditions they assume at the dislocation edge are not correct.

The purpose of this work is to examine critically the boundary conditions at a lineage line and to solve the continuity equation in a simple one-dimensional case utilizing the correct boundary conditions. The results of this analysis are more general than those of Vogel, Read, and Lovell, since the effect of a constant field in the $+x$ direction on the solution of the problem is calculated. It will be necessary in treating this problem to investigate the gross fluxes of carriers across the lineage plane as well as the net flux given by diffusion theory. Thus it will be necessary to invoke kinetic theory to determine the actual magnitudes of the various flux components which are to be considered.

II. BOUNDARY CONDITIONS AT A LINEAGE BOUNDARY

The boundary conditions on the continuity equation for injected carriers at a lineage boundary will be developed via the method of multiple reflections which was introduced in a previous paper.[4] Consider a net flow of A injected carriers per cm^2 per sec to enter the neighborhood of a lineage plane.[5] Then, if the "transmission coefficient" of the array of edge-type dislocations (the probability that a carrier will not recombine in an encounter with the array) is denoted by T, and the probability that a carrier starting out at the lineage boundary into the bulk material to the left or right of the edge will return to the boundary before recombining in the bulk is denoted by "bulk reflection coefficients" B' and B to the left and right of the lineage boundary, respectively, the situation is as represented in Fig. 1. We assume that there is a constant field E_0 sweeping the carriers to the right along OX, giving the carriers a drift velocity $v_0=\mu E_0$ in that direction, where μ is the drift mobility of the injected carriers.

It can be shown by integrating fluxes over a Maxwell-Boltzmann velocity distribution upon which has been superimposed the small drift velocity v_0 that the number of carriers traversing unit area oriented normal to the x-axis per unit time in the presence of the field E_0 along the x-axis is

$$\nu=N(\tfrac{1}{4}\bar{c}+\tfrac{1}{2}v_0), \tag{1}$$

where N represents the number of carriers per unit volume,[6] and where the mean thermal velocity $\bar{c}$ is

[1] A. H. Cottrell, Progr. Metal Phys. **1**, 78 (1950).
[2] F. L. Vogel *et al.*, Phys. Rev. **90**, 489 (1953).
[3] Vogel, Read, and Lovell, Phys. Rev. **94**, 1791 (1954).
[4] J. P. McKelvey and R. L. Longini, J. Appl. Phys. **25**, 634 (1954).
[5] One assumes here that there is a plane source of injected minority carriers to the left of and parallel to the lineage plane. This assumption gives rise to a certain asymmetry in the results which are derived. The assumption of plane sources of injected carriers to the right and to the left of the lineage plane would yield symmetrical relations. This case will not be worked out in detail here, since the simpler case is adequate to predict all the experimental results which are considered, and since the extension of the simpler case to the symmetric case is obvious.
[6] This result is just what one would expect from a naive consideration of the action of a small field on an ensemble of independent particles with mean velocity $\bar{c}$. The fluxes to the right and to the left across unit area per unit time are $N\bar{c}/4$ in the absence of the field. With the field on, the flux to the right increases by a small amount and the flux to the left decreases by the same small amount. But the difference of the fluxes must be Nv_0; thus the fluxes to the right and to the left in the presence of field should be $N(\frac{1}{4}\bar{c}+\frac{1}{2}v_0)$ and $N(\frac{1}{4}\bar{c}-\frac{1}{2}v_0)$, respectively.

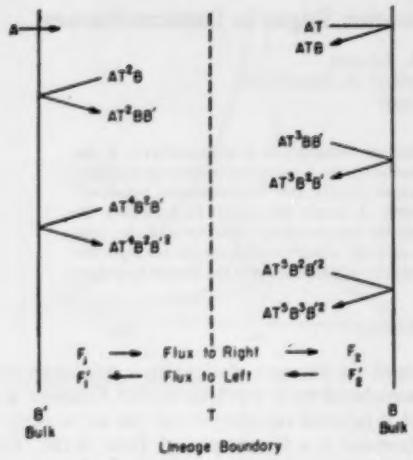


FIG. 1. Multiple reflection representation of carrier fluxes in the neighborhood of a lineage boundary.

given by

$$\bar{c}=(8kT/\pi m)^{\frac{1}{2}}. \tag{2}$$

The assumption of Maxwell-Boltzmann statistics is justified if (1) the number of carriers is small compared to the number of available states so that the carriers may be treated as a nondegenerate electron gas, (2) the drift velocity v_0 is small compared to the mean thermal velocity $\bar{c}$, so that the departure from equilibrium conditions is small, and (3) the surfaces of constant energy in **k**-space are spherical. These assumptions are justifiable for holes in silicon and germanium of relatively high purity at moderately high temperatures, unless very strong fields are applied.

One can now investigate the carrier fluxes at a lineage plane. From Fig. 1, one can write, summing the geometric series which arise on adding the reflected flux components to get the total fluxes,

$$F_1=A\sum_{n=0}^{\infty}T^{2n}B^nB'^n=\frac{A}{1-T^2BB'};$$

similarly

$$F_1'=\frac{AT^2B}{1-T^2BB'},\quad F_2=\frac{AT}{1-T^2BB'},\quad F_2'=\frac{ATB}{1-T^2BB'}. \tag{3}$$

The total fluxes to the right and to the left can now be calculated from (1) and (3). It is necessary, however, to assume that the fluxes to the right F_1 and F_2 originate at a virtual source situated to the left of the lineage boundary a distance $\alpha\lambda$, where λ is the carrier mean free path and α is a numerical constant of the order of unity. The necessity of this assumption is clear when one considers the case where $T=1$. In this case, no concentration gradient at the boundary is obtained even when a carrier source to the left of the boundary is present, unless the virtual source mentioned above is assumed to be present. This can be seen to be true by inspection of the equations for the net fluxes across the boundary which will be written in connection with the present derivation. Similarly, the fluxes F_1' and F_2' are assumed to originate at a virtual source situated a distance $\alpha\lambda$ to the right of the lineage boundary. If the fractional concentration variation over a mean free path is small, and if $\mu E_0 \ll \bar{c}$, the factor α will be a constant independent of concentration, concentration gradient and applied field. From (1) and (3), then, noting the relative directions of the drift velocity and the fluxes from Fig. 1, and denoting the injected carrier concentration *at the lineage boundary* as p_1 and p_2 in the left-hand region and the right-hand region, respectively,

$$F_1+F_1'=\left(\frac{\bar{c}}{4}+\frac{\mu E_0}{2}\right)\left(p_1-\alpha\lambda\frac{\partial p_1}{\partial x}\right)$$

$$+T\left(\frac{\bar{c}}{4}-\frac{\mu E_0}{2}\right)\left(p_2+\alpha\lambda\frac{\partial p_2}{\partial x}\right)=\frac{A(1+T^2B)}{1-T^2BB'}, \tag{4a}$$

$$F_2+F_2'=T\left(\frac{\bar{c}}{4}+\frac{\mu E_0}{2}\right)\left(p_1-\alpha\lambda\frac{\partial p_1}{\partial x}\right)$$

$$+\left(\frac{\bar{c}}{4}-\frac{\mu E_0}{2}\right)\left(p_2+\alpha\lambda\frac{\partial p_2}{\partial x}\right)=\frac{AT(1+B)}{1-T^2BB'}. \tag{4b}$$

In the above equations, and in the subsequent development, p_1 and p_2 refer to the concentration of *injected* carriers, that is, to the actual carrier concentration minus the equilibrium value of carrier concentration. If the injected carrier concentration is small compared to the total concentration of charge carriers, the distribution in energy and in velocity of these carriers will be very nearly that which was assumed in deriving (1). The flux A in Fig. 1 is to be interpreted as a flux of injected carriers only; any flux to the right in this diagram due to the presence of the equilibrium carrier concentration will be balanced by an equal flux to the left due to the equilibrium carrier concentration. This is true because the principle of detailed balancing requires that at equilibrium the rate of recombination of carriers be just equal to the rate of thermal generation of pairs, so that any flux of equilibrium carriers due to recombination at the lineage boundary must be accompanied by an equal and opposite flux of equilibrium carriers due to thermal generation at the boundary. Thus, only the net flux of injected carriers need be considered, provided p_1 and p_2 are thought of as injected carrier densities.

Multiplying (4b) by T and subtracting from (4a), one can solve for A:

$$A=(\tfrac{1}{4}\bar{c}+\tfrac{1}{2}\mu E_0)(p_1-\alpha\lambda\partial p_1/\partial x)(1-T^2BB'). \tag{5}$$

Substituting (5) into (4a), it is found that

$$\frac{p_2+\alpha\lambda\partial p_2/\partial x}{p_1-\alpha\lambda\partial p_1/\partial x}=TB\frac{\frac{1}{2}\mathcal{E}+\mu E_0}{\frac{1}{2}\mathcal{E}-\mu E_0}. \tag{6}$$

If one calculates F_1-F_1' in a similar manner, the result may be equated to $-D(\partial p_1/\partial x)+\mu E_0 p_1$ at the boundary, since both these expressions give the net flow of carriers out of the left hand region of Fig. 1. The same argument holds regarding the flow of carriers into the region to the right of the lineage line. Thus

$$-D\frac{\partial p_2}{\partial x}+\mu E_0 p_2=F_2-F_2'=T\left(\frac{\mathcal{E}}{4}+\frac{\mu E_0}{2}\right)\left(p_1-\alpha\lambda\frac{\partial p_1}{\partial x}\right)-\left(\frac{\mathcal{E}}{4}-\frac{\mu E_0}{2}\right)\left(p_2+\alpha\lambda\frac{\partial p_2}{\partial x}\right), \tag{7a}$$

$$-D\frac{\partial p_1}{\partial x}+\mu E_0 p_1=F_1-F_1'=\left(\frac{\mathcal{E}}{4}+\frac{\mu E_0}{2}\right)\left(p_1-\alpha\lambda\frac{\partial p_1}{\partial x}\right)-T\left(\frac{\mathcal{E}}{4}-\frac{\mu E_0}{2}\right)\left(p_2+\alpha\lambda\frac{\partial p_2}{\partial x}\right). \tag{7b}$$

If $T=1$, $p_1=p_2=p$ and both equations reduce to the same form. In this case, both of the above equations can be written

$$-D\frac{\partial p}{\partial x}=-\mu E_0 p+\left(\frac{\mathcal{E}}{4}+\frac{\mu E_0}{2}\right)\left(p-\alpha\lambda\frac{\partial p}{\partial x}\right)-\left(\frac{\mathcal{E}}{4}-\frac{\mu E_0}{2}\right)\left(p+\alpha\lambda\frac{\partial p}{\partial x}\right)=-\frac{\alpha\lambda\mathcal{E}}{2}\frac{\partial p}{\partial x}.$$

From kinetic theory one can show that $D=\lambda\mathcal{E}/3$. Substituting this result in the above equation and solving for α gives the result[7]

$$\alpha=\tfrac{2}{3}. \tag{8}$$

If (6) is substituted into (7a) and all quantities expressed in terms of p_2 and $\partial p_2/\partial x$, the quantities containing $\partial p_2/\partial x$ may be grouped on the left-hand side of the resulting equation, and all quantities containing p_2 may be grouped on the right-hand side, allowing the boundary condition to be expressed

$$-D\partial p_2/\partial x=s_2 p_2, \tag{9a}$$

where

$$s_2=\left[\frac{1}{2}\left(\frac{\mathcal{E}}{2}+\mu E_0\right)(1-B)-\mu E_0\right]\Big/\left[1-\frac{1}{\mathcal{E}}\left(\frac{\mathcal{E}}{2}+\mu E_0\right)(1-B)\right]. \tag{9b}$$

[7] This same result can be obtained by averaging the projection of the mean free path on the normal to the lineage plane over all possible orientations of the carrier path with respect to this normal for net carrier fluxes along the positive x-direction. The result is, of course, well known.

The relations $D=\lambda\mathcal{E}/3$ and $\alpha=\frac{2}{3}$ have been utilized to arrive at (9b). Similarly, (6) can be substituted into (7b), and, expressing all quantities in terms of p_1 and $\partial p_1/\partial x$, the result is

$$-D\partial p_1/\partial x=s_1 p_1, \tag{10a}$$

where

$$s_1=\left[\frac{1}{2}\left(\frac{\mathcal{E}}{2}+\mu E_0\right)(1-T^2B)-\mu E_0\right]\Big/\left[1-\frac{1}{\mathcal{E}}\left(\frac{\mathcal{E}}{2}+\mu E_0\right)(1-T^2B)\right]. \tag{10b}$$

Equations (6), (9), and (10) form a set of boundary conditions on the continuity equation which are applicable to recombination at a lineage boundary under the rather modest assumptions that $\mu E_0 \ll \mathcal{E}/2$ and the fractional variation in carrier concentration over a mean free path is small. In these equations, if $T=1$, it can easily be shown that $s_1=s_2$. If $T<1$, $s_1>s_2$. In (9b) and (10b) above one might be tempted to approximate $\frac{1}{2}\mathcal{E}+\mu E_0$ and $\frac{1}{2}\mathcal{E}-\mu E_0$ by $\frac{1}{2}\mathcal{E}$, neglecting the μE_0 in comparison with $\frac{1}{2}\mathcal{E}$. It happens, however, that the expressions derived by subsequent manipulation involving these equations are simpler when the exact expressions are used and this approximation not employed. Of course, it is impossible to neglect μE_0 compared with $\frac{1}{2}(\frac{1}{2}\mathcal{E}+\mu E_0)(1-B)$ and $\frac{1}{2}(\frac{1}{2}\mathcal{E}+\mu E_0)\times(1-T^2B)$ in the numerators of (9b) and (10b), respectively, since in many cases T and B are very close to unity and these terms might well be of the same order of magnitude as μE_0.

In Eq. (10b), s_1 is expressed in terms of T and B. One wishes to eliminate B from (10b) and express T as a function of s_1 and s_2. This can be accomplished quite directly. Solving (9b) for B, the result is

$$B=\frac{\frac{1}{2}\mathcal{E}-\mu E_0}{\frac{1}{2}\mathcal{E}+\mu E_0}\cdot\frac{1-2s_2/\mathcal{E}}{1+2s_2/\mathcal{E}}. \tag{11}$$

Now, Eq. (10b) can be solved for T^2B, and the value of B found in (11) substituted in the resulting equation. One finds

$$T^2=\frac{1-2s_1/\mathcal{E}}{1+2s_1/\mathcal{E}}\cdot\frac{1+2s_2/\mathcal{E}}{1-2s_2/\mathcal{E}}. \tag{12}$$

Equations (11) and (12) accomplish the desired result. B involves s_2 only, whereas T is a function of both s_1 and s_2.

The discontinuous nature of the solutions and their derivatives can be easily exhibited from the above equations. If (9a) and (10a) are substituted into (6), the relation $D=\lambda\mathcal{E}/3$ utilized, and B eliminated by (11), one can show that at the boundary

$$\frac{p_2}{p_1}=T\cdot\frac{1+2s_1/\mathcal{E}}{1+2s_2/\mathcal{E}}. \tag{13}$$

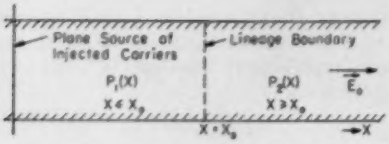


FIG. 2. Idealized geometry to which the boundary conditions must be applied to solve the continuity equation for the suggested experiment.

Using (13) in (9a) and (10a) it is also obvious that

$$\frac{\partial p_2/\partial x}{\partial p_1/\partial x}=T\left(\frac{s_2}{s_1}\right)\cdot\frac{1+2s_1/\mathcal{E}}{1+2s_2/\mathcal{E}}. \tag{14}$$

The expressions (13) and (14) do not reduce to unity for T different from unity; therefore both the concentration and concentration gradient are discontinuous at the boundary. In all cases s_1 and s_2 are much smaller than $\mathcal{E}$, and thus the value of p_2/p_1 at the boundary is approximately T, and the value of $(\partial p_2/\partial x)/(\partial p_1/\partial x)$ is approximately Ts_2/s_1.

The continuity equation for injected carriers is a second-order differential equation, for which there are two different solutions in the regions to the left and right of the lineage boundary in Fig. 1. Four boundary conditions are therefore required to establish the solutions in both regions. Equations (6), (9), and (10) must hold at the lineage boundary, and these equations constitute three of the boundary conditions. In a situation such as that shown in Fig. 1, it is obvious that if $p_3(x)$ represents the injected carrier concentration in the region to the right of the lineage boundary that

$$\lim_{x\to\infty} p_3(x)=0. \tag{15}$$

This condition represents the fourth boundary condition on the continuity equation.

Note that although Eqs. (13) and (14) indicate discontinuities in the injected carrier concentration and its derivative at the lineage boundary, these discontinuities are dependent on the assumption that the lineage boundary can be regarded as a mathematical plane. If one could examine the value of the injected carrier concentration as a function of distance along a path normal to the lineage plane but not intersecting a dislocation edge, one would find a rapid but continuous change in this quantity (and its derivative) in the neighborhood of the idealized dislocation plane. In order to make the calculations, one assumes recombination at a mathematical plane intersecting the dislocation edges and for this model the discontinuities are found. The concentrations and gradients given by the idealized treatment are not significantly different from the actual values if the distance from the lineage plane is more than a few microns.

III. EXPERIMENTAL DETERMINATION OF TRANSMISSION COEFFICIENT

By using the boundary conditions developed in the previous section, the continuity equation for injected carriers can be solved quite readily for the geometry shown in Fig. 2. The constant field E_0 is applied so as always to sweep carriers to the right. The solution of the continuity equation in this geometry, the details of which will not be presented here, can be utilized to determine the transmission coefficient T with the aid of an experiment which approximates the geometry of Fig. 2. Such an experiment is shown in Fig. 3. The experimental arrangement shown will enable one to determine T if the solution of the continuity equation is known, and also to determine the value of the minority carrier diffusion length L and the drift mobility μ, which must be known if one wishes to solve for T.

First of all, the carrier concentration can be measured by measuring the collector currents at A, B, and C, relative carrier concentrations being obtained by calibrating each collector with respect to its reverse saturation current in the absence of injected carriers. Then L can be determined at once, since

$$e^{-b/L}=\frac{p_2(a+b)}{p_2(a)}=\frac{(I_C/I_{0C})-1}{(I_B/I_{0B})-1}=\frac{I_{0B}}{I_{0C}}\frac{(I_C-I_{0C})}{(I_B-I_{0B})}. \tag{16}$$

Here I_C refers to the collector current at C and I_{0C} refers to the reverse saturation current at C in the absence of injected carriers. Similar notation is used with reference to electrodes A and B. The distance a can be measured with a measuring microscope. The drift mobility μ can be measured by means of the Haynes-Shockley technique[8] using electrode B as an emitter and electrode C as a collector. In actual practice this measurement would be made first, so that the unformed electrode B could be used as an emitter, and the measurement of L would then be made, electrode B having been formed for efficient collector operation. The diffusion constant D is obtained from the drift

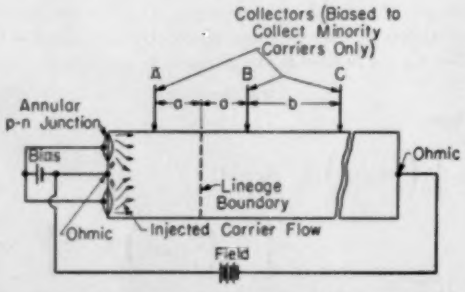


FIG. 3. Schematic diagram of the experimental apparatus required to measure transmission coefficient.

[8] J. R. Haynes and W. Shockley, Phys. Rev. 81, 835 (1951).

mobility μ by the well-known Einstein relation. T can be found by finding the ratio $p_1(-a)/p_2(a) \equiv R$ and employing the solution of the continuity equation.

In practice some modification of the foregoing experimental techniques might be desirable. For instance, the use of pulsed emitter voltages and ac detectors might be preferable to the dc circuit shown. Some care must naturally be taken to find suitable samples for measurements. It is desirable to have a sample with one isolated lineage plane traversing it normal to its axis. Randomly scattered dislocations throughout the rest of the sample would not affect the transmission coefficient measurement, since the effect of these would be incorporated into the parameter L. Other lineage planes, if present, would have to be several diffusion lengths away from the one under study. High bulk lifetimes are desirable, so that probe spacings may be relatively large and signal ratios due to bulk recombination small.

Surface recombination will affect the measurement of L, but only to the extent that the measured decay constant will be an "effective" diffusion length which is partly due to bulk recombination, and partly due to surface recombination. However, this effective diffusion length also appears in the solution of the continuity equation.

IV. RELATION BETWEEN THE TRANSMISSION COEFFICIENT AND THE RECOMBINATION CROSS SECTION

If a flow of carriers is incident on a regular array of recombination centers, as in Fig. 4, the relation between the probability of transmission T and the recombination cross section can be obtained by averaging over the angle of incidence. Suppose that the lineage plane forms a regular array of recombination centers having cylindrical symmetry about the terminations of the lattice planes which form the dislocations. For a given angle of incidence, θ, carriers which are initially in the shaded regions will recombine. Obviously, if $|\theta| > |\theta_0|$, where

$$\cos\theta_0 = g/h, \tag{17}$$

there will be no transmission at all. Here g is the diameter of the cylindrical recombination area and h is the spacing between dislocations. The transmission coefficient T is merely the average of the ratio of shaded area to total area over the angle θ between the limits of $\pi/2$ to $-\pi/2$ with the proviso that this ratio is zero for $\theta_0 < \theta < \pi/2$ and $-\pi/2 < \theta < -\theta_0$. Therefore,

$$T = \frac{1}{\pi}\int_{-\theta_0}^{\theta_0} \frac{h\cos\theta - g}{h\cos\theta}\,d\theta = \frac{2}{\pi}\int_0^{\theta_0}\left(1 - \frac{h}{g}\sec\theta\right)d\theta. \tag{18}$$

After integrating, substituting the limits of integration, and expressing all trigonometric functions of θ_0 in terms of g and h by (17), the result is

$$T = \frac{2}{\pi}\cos^{-1}\frac{g}{h} - \frac{2g}{\pi h}\ln\left\{\frac{h}{g}\left[1 + \left(1 - \frac{g^2}{h^2}\right)^{\frac{1}{2}}\right]\right\}. \tag{19}$$

From Fig. 4, it is clear that g represents the recombination cross section per unit length of dislocation for a single edge-type dislocation in the array. Equation (19) therefore represents the desired relation between transmission coefficient and recombination cross section. Figure 5 shows a plot of (19). If $g=0$, $T=1$ as expected. For $g>0$, T decreases as shown in the plot.

It has been shown that a measurement of the trans-

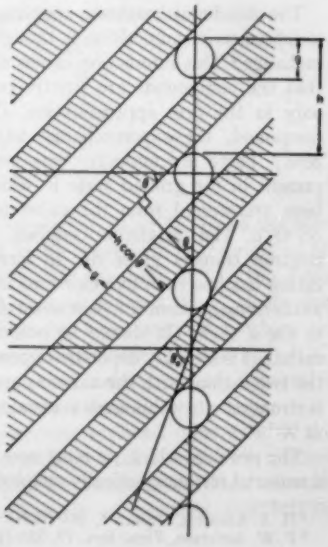


FIG. 4. Transmission and absorption of carriers obliquely incident on a regular array of cylindrical recombination centers representing a lineage boundary.

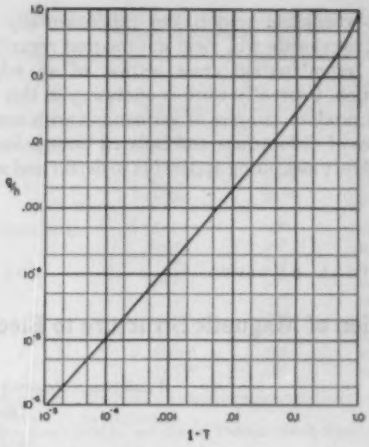


FIG. 5. Probability 1-T of recombination on collision with a lineage boundary vs recombination cross section per unit area of lineage plane, g/h.

mission coefficient undertaken experimentally as described previously will yield information regarding the actual recombination cross section of an edge-type dislocation. Currently work is underway in this laboratory to produce samples of germanium with controlled densities of dislocations and isolated lineage lines, and to construct measuring apparatus to detect and measure the effects described in this paper. This work will be discussed in a future publication.

ACKNOWLEDGMENTS

The authors wish to acknowledge the very helpful suggestions and discussions which Dr. F. Keffer of these laboratories has contributed to the present work.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Relation of Magnetic Structure to Electrical Conductivity in NiO and Related Compounds*

R. R. Heikes
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received October 1, 1954)

A qualitative theory is presented in an effort to understand the insulating properties of NiO. If one generalizes this theory to include all stoichiometric binary compounds, having a transition element as cation and an element from group V or VI as anion, a rather startling correlation between the magnetic and electric properties is predicted. Ferromagnetic compounds should be good conductors with a positive temperature coefficient of resistivity while antiferromagnets should be insulators with a negative temperature coefficient of resistivity. The available data indicate that this correlation does exist.

I. INTRODUCTION

NICKEL oxide has been the subject of much discussion because of its electrical properties. As the Ni ions form an incomplete d band, NiO should, by the usual criterion, be a conductor. However, it is found to be an insulator. This same anomaly is also possessed by a number of other materials: CuO, Fe_2O_3, MnO, MnS, MnSe, CoO, to name a few. Verwey and de Boer[1] have attempted to explain this group of materials on a band model by assuming that the potential barrier for electron motion is sufficiently high to prevent conduction. However, Mott[2] points out that if this were the case one would not be justified in using a band approach. Instead Mott treats the problem using a Heitler-London model which requires the presence of mixed valence states for conductivity. He suggests that conductivity might take place by the formation of Ni^+-Ni^{+++} ion pairs if enough energy were available to break up these ion pairs or else if enough pairs were formed so that many degenerate positions existed for the Ni^+ and Ni^{+++} ions. He makes the further tentative suggestion that the lowest energy state will not contain sufficient Ni^+-Ni^{+++} ion pairs. Thus, NiO should be, as is the case, an insulator at ordinary temperatures.

The suggestion of Mott that the electrons are too tightly bound to admit conductivity does not seem tenable when one considers the magnetic properties of the crystal. All the aforementioned examples are found to be antiferromagnetic. This antiferromagnetism is presumed to result from the superexchange interaction, which gives rise to configurations of the type Ni^+-Ni^{++} rather than Ni^+-Ni^{+++} as suggested by Mott, since in this case the electron transfer is actually from O^{--} to Ni^{++} as shown in Sec. III below. However, the essential idea of Mott's argument, that the metal ions must exist in different valence states, is still valid.

II. SUPEREXCHANGE

The dominant magnetic coupling in compounds of stoichiometric composition is thought to be the superexchange.[3] This interaction comes about from the fact that the compounds are strictly polar (i.e., $Ni^{++}O^{=}$) only in the first approximation. (In a perfect ionic compound, the superexchange interaction would be zero.) There are actually "less ionic" configurations present in the ground state in which an electron has been transferred from the anion to the cation (e.g., Ni^+O^-). The magnetic coupling arising from this electron transfer gives rise to ferromagnetism if the cation has less than five electrons in the d shell and to antiferromagnetism if five or more electrons are present in the d shell. (It should be pointed out that superexchange is strongly dependent upon the angle made by the two cations with the anion as apex. The interaction is strongest when this angle is equal to 180° and weakest at 90°.)[4]

The problem which we must now understand is why a material ferromagnetically coupled by superexchange

* Work supported in part by Wright Air Development Center, U. S. Air Force.
[1] J. H. de Boer, and E. J. W. Verwey, Proc. Phys. Soc. (London) **49**, 59 (1937).
[2] N. F. Mott, Proc. Phys. Soc. (London) **A62**, 416 (1949).
[3] H. A. Kramers, Physica **1**, 183 (1934).
[4] P. W. Anderson, Phys. Rev. **79**, 350 (1950).

is a conductor while those coupled antiferromagnetically are insulators.

III. RELATION TO CONDUCTIVITY

The fact that these "less ionic" states are present a fraction of the time leads immediately to a model for conduction requiring no activation energy. One might thus assume that conduction in materials magnetically coupled by superexchange would be metallic (positive temperature coefficient of resistance) in nature. However, at this point one must consider the difference between antiferromagnetic and ferromagnetic superexchange coupling.

Let us consider first the ferromagnetic case (less than half-filled d shell) which is represented schematically below. (The ionic state is shown.)

M^{++}	$O^{=}$	M^{++}	$O^{=}$	M^{++}
↑		↑		↑
↑	↑ ↓	↑	↓ ↑	↑

As we have already mentioned, the transition from the ionic configuration to the less ionic one is effected by the movement of an electron from the $O^{=}$ to a neighboring M^{++}. Under the influence of an applied electric field, one of the $O^{=}$ p electrons can travel down the line of atoms represented above, the only requirement being that the vacancies be present in the p shells of the appropriate $O^{=}$ ions. Furthermore, from a knowledge of superexchange it is known that only the oxygen p electrons with spin parallel to that of the metal d shells engage in these transitions. Thus, in the schematic diagram only electrons with up spin will take part in conduction. (Transitions involving the electron with down spin are not forbidden but are energetically much less favorable.)

Now let us consider the antiferromagnetic case.

M^{++}	$O^{=}$	M^{++}	$O^{=}$	M^{++}
↑		↓		↑
↑		↓		↑
↑		↓		↑
↑	↓ ↑ 1	↓	↓ ↑	↑
↑		↓		↑
I		II		III

Electron 1 can transfer to atom II. However, even if a vacancy exists in the oxygen ion to the right of II, electron 1 cannot go on to atom III as the Pauli principle would be thereby violated. Thus, conduction by means of the deviation from perfect ionic character is forbidden in an antiferromagnetic lattice at 0°K, except to the extent that metal ion configurations of higher principal quantum number exist.

At a finite temperature reversed spins exist in the otherwise antiferromagnetic lattice. Each such reversal presents a possibility for conduction. Thus, an antiferromagnet should have a negative temperature coefficient of resistivity.

TABLE I Magnetic and electrical properties of binary compounds

	Substance	$-\sigma_{298}$ (room temp.)	Temperature coefficient of resistivity
Ferromagnetic	MnBi[a]	10^{-4}	positive
	MnAs[a]	10^{-4}	positive
	MnSb[a]	10^{-5}	positive
	CrTe[b]	10^{-4}	positive
Antiferromagnetic	MnO[c]	10^{9}	negative
	FeO[c]	10^{8}	negative
	CoO[c]	10^{8}	negative
	NiO[c]	10^{8}	negative
	CuO[c]	10^{10}	negative
	Fe_2O_3[c]	10^{10}	negative
	MnS[d]	10^{6}	negative
	MnSe[d]	10^{6}	negative
	MnTe[d]	10^{9}	negative

[a] R. R. Heikes (to be published).
[b] Fakidov, Grazhdankina, and Kikoin, Doklady Akad. Nauk. (S.S.S.R.) 68, 491 (1949).
[c] See reference 1.
[d] C. F. Squire, Phys. Rev. 56, 960 (1943).

IV. NiO AND OTHER ANTIFERROMAGNETS

Let us now see how NiO can be understood from the foregoing analysis. Remembering that superexchange is only appreciable when the metal ion-oxygen-metal ion angle is 180°, we can think of the Ni ions in NiO as a face-centered cubic lattice made up of four essentially noninteracting simple cubic lattices, each antiferromagnetic within itself. On this basis, each Ni ion interacts with six Ni ions (next nearest neighbors in the actual crystal) all of opposite spin orientation. Thus, NiO should be nonconducting.

It should be noted that although a Ni ion has no neighbors of parallel spin at 180°, there are some at 90°. The present model requires this superexchange coupling to be negligible. From the present theory of superexchange it is clear that the 90° coupling will be small; however, it is not clear that it will be negligible.

A possible objection to the present model may be mentioned here. One might expect that at T_c large anomalies in conductivity would be found and that above T_c conductivity should be large. At present there are not sufficient data on this subject. However, it is worth mentioning that the above effects probably will not occur in most antiferromagnetics because of the pronounced short range order which exists even up to 2 and $3T_c$. Such short range order on the present model would be sufficient to hinder conductivity.

Table I cites the resistance and its temperature dependence for all antiferromagnetic compounds on which data are available. (Resistivities are given only to an order of magnitude.) It should be noted that all have a high resistivity and a negative temperature coefficient as is expected.[5]

[5] MnTe, which has a somewhat anomalous temperature dependence of the resistivity will be discussed in a succeeding paper.

V. FERROMAGNETIC MATERIALS

Table I also lists the resistivities of the ferromagnetic stoichiometric compounds. It is to be noted that these substances have resistivities 5 to 14 orders of magnitude lower. In addition, the conduction is metallic in sign. In order to see that the order of magnitude of the resistivity is correct, let us make a calculation on a simple model. Considering (see Zener's[6] double exchange) conductivity to be a diffusion process, Zener showed that one could write the following relation:

$$\sigma = Ne^2a^2J/kTh,$$

where N is the number of electrons per unit volume participating in the diffusion process (e.g., equal to the number of Ni^+ ions in NiO), a is the lattice constant, and J is the integral representing the transition of an electron from the anion to the cation. Taking Van Vleck's[7] estimate of 10^3 cm^{-1} for J and assuming that $N \simeq 0.1 \times$ number of anions, one finds $10^3\Omega^{-1}$ cm^{-1}. This is the correct order of magnitude for the ferromagnetic materials.

[6] C. Zener, Phys. Rev. 82, 403 (1951).

SUMMARY

The present model treats NiO on an atomic basis, the lack of conduction arising from the action of the Pauli exclusion principle. This treatment should be valid as long as cation states of principle quantum number greater than three can be neglected.

The data in Table I show that the predicted qualitative correlation between the magnetic and electric properties is borne out.

[7] J. H. Van Vleck, Grenoble Conference 114 (1951).

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Avalanche Breakdown in Germanium

S. L. MILLER
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received May 11, 1955)

It is shown that all germanium junctions studied break down as the result of the same avalanche process found in silicon. An empirical expression for the multiplication inherent in this breakdown process is given for step junctions. Ionization rates for holes and electrons in Ge are derived with the use of this expression. The ionization rate for holes is larger than that for electrons by about a factor of two. The agreement between these ionization rates as a function of field and the theory of Wolff is excellent. It is determined that the threshold for electron-hole pair production is about 1.50 ev and the mean free path for electron (or hole)-phonon collisions is about 130 A.

INTRODUCTION

WHEN pn junctions are reverse-biased to sufficiently high voltage, a breakdown occurs and large currents begin to flow. The principal theories for this breakdown depend on internal field emission or a solid state analogue of the Townsend β avalanche breakdown in gases.[1,2] The former mechanism was first proposed by Zener[3] for dielectric breakdown. It involves the direct excitation of electrons from the valence to the conduction bands in high electric fields. In the latter mechanism, electrons or holes in high fields interact with valence electrons to produce electron-hole pairs. An important difference between the two mechanisms is that in the latter case multiplication of charge injected into the junction takes place at voltages below the breakdown voltage. In fact, breakdown is defined as the point at which this multiplication becomes very large. The Zener theory has no such multiplicative effects associated with it.

Recently, McKay and McAfee[4] have demonstrated that charge multiplication takes place in some germanium and silicon junctions at prebreakdown voltages. McKay[2] has found that all silicon junctions in the range studied have multiplicative breakdowns. Breakdowns possibly attributable to internal field emission have been reported in some narrow germanium junctions.[4]

The measurements described in this paper were undertaken to understand the breakdown mechanism in germanium junctions. However, the analysis developed should be applicable to other semiconductors. As a result of the measurements, there is strong evidence that the breakdown process in germanium is the avalanche process and that internal field emission has not been observed. It has also been possible to differentiate between the roles of holes and electrons in the avalanche breakdown process by an extension of the Townsend β theory and to compare the ionization rates obtained with the recent theory of Wolff.[5]

[1] K. G. McKay and K. B. McAfee, Phys. Rev. 91, 1079 (1953).
[2] K. G. McKay, Phys. Rev. 94, 877 (1954).
[3] C. Zener, Proc. Roy. Soc. (London) 145, 523 (1934).
[4] McAfee, Ryder, Shockley, and Sparks, Phys. Rev. 83, 650 (1951).
[5] P. A. Wolff, Phys. Rev. 95, 1415 (1954).

THEORY

In the field emission hypothesis, breakdown occurs sharply when a certain critical field is reached anywhere in the junction. This critical field is a constant in a given semiconductor at a given temperature. In a step junction, where the conductivity type varies abruptly from p to n and the impurity concentration is very much greater on one side than the other, the maximum electric field is given by[6]

$$E_M = 2V/W = 2V/W_1 V^{\frac{1}{2}} = 2V^{\frac{1}{2}}/W_1 = KV^{\frac{1}{2}}(|N_D - N_A|)^{\frac{1}{4}}, \quad (1)$$

where W = width of the space charge region, W_1 = width constant of the junction $= (1.77\times10^7/|N_D-N_A|)^{\frac{1}{2}}$ for Ge, $|N_D-N_A| = N_I$ = the net impurity concentration on the high resistivity side of the junction, and V = the voltage applied across the junction plus the built-in voltage of the junction.

Thus the critical electric field would be reached when $V = V_B$, the breakdown voltage. Then

$$(E_M)_{\text{critical}} = \text{a constant} = KV_B^{\frac{1}{2}}(|N_D-N_A|)^{\frac{1}{4}},$$

or

$$V_B = \text{constant}/|N_D - N_A|. \quad (2)$$

Hence, the Zener theory results in a breakdown voltage which is inversely proportional to the net impurity concentration on the high-resistivity side for this type of junction. In the range of resistivities where mobility is essentially constant, this means that V_B would be proportional to resistivity. This relation does not hold even for the narrowest germanium junctions investigated.

McKay has applied a modified form of the Townsend β discharge theory for gases to the multiplicative breakdown process in silicon junctions. Under the assumption that electrons and holes have equal ionization rates he obtained an expression for the ionization rate at the maximum field in the junction, $\alpha_i(E_M)$. In step junctions, like those discussed above, numerical values for $\alpha_i(E_M)$ could then be determined from measurements of the multiplication, M, of a junction *vs* E_M or alternatively from measurements of E_M at $M=\infty$ or breakdown *vs* W_1 for various junctions. The multiplication experiments were performed on diodes with the carriers injected by α-particle bombardment.

It has been observed experimentally in this research that the multiplication of minority carriers coming from the high-resistivity side of germanium step junctions closely follows the empirical expression:

$$M(V) = 1/[1-(V/V_B)^n], \quad (3)$$

where V_B is the body breakdown of the junction and the parameter n is a number which depends on the resistivity and resistivity type of the high-resistivity side of the junction. It will be shown below that n is quite different for an np^+ and a pn^+ step junction with the same impurity concentration on the high-resistivity side for germanium. This experimental fact can only be interpreted as proof that the ionization rate for electrons is different from that of holes. Furthermore, the existence of a good analytic approximation to the variation of multiplication with voltage makes possible the derivation of analytical expressions for the ionization rates themselves.

$\alpha_i(E)$ is now defined as the number of electron-hole pairs produced by an electron per centimeter travelled in the direction of the field, E. $\beta_i(E)$ is the analogous quantity for holes. Following the notation of McKay, n_0 is the number of electrons entering the junction at $x=0$, n_1 is the number of electrons produced by electrons or holes between 0 and x, and n_2 is the number of electrons produced between x and W. Then the number of electrons produced between x and $x+dx$ is

$$dn_1 = (n_0+n_1)(\alpha_i-\beta_i)dx + (n_0+n_1+n_2)\beta_i dx.$$

This is integrated with the boundary conditions $n_1=0$ at $x=0$ and $n_2=0$ at $x=W$. We then obtain

$$1-\frac{1}{M} = \int_0^W \alpha_i \exp\left[-\int_0^x (\alpha_i-\beta_i)dx'\right]dx, \quad (4)$$

where $M = (n_0+n_1+n_2)/n_0$ = the multiplication factor.

Breakdown occurs or $M\to\infty$ when

$$\int_0^W \alpha_i \exp\left[-\int_0^x (\alpha_i-\beta_i)dx'\right]dx = 1. \quad (5)$$

It should be noted that in the case of those junctions in which the injected carriers are holes instead of electrons, equations (4) and (5) hold with the α_i's and β_i's reversed. When $\alpha_i=\beta_i$ these equations reduce to McKay's result. The assumptions implicit in them are the same as those of McKay:

(1) The ionization rates of both holes and electrons are only functions of the electric field.

(2) The loss of carriers in the junction by recombination is negligible.

(3) Mutual interactions between carriers are negligible.

(4) The density of carriers in the junction is small enough so that there is no change in the field configuration in the junction. That is, there are no space-charge effects.

In the case of a step junction like that already discussed, essentially all of the space-charge region is on the high-resistivity side and the electric field distribution is given by

$$E = E_M x/W, \quad (6)$$

where $W-x$ is the distance from the junction. This distribution is shown in Fig. 1. The $x=0$ point has been chosen as the point where $E=0$ since the particles

[6] The reader is referred to W. Shockley, Bell System Tech. J. 28, 435 (1949) for a detailed treatment of p-n junction theory.

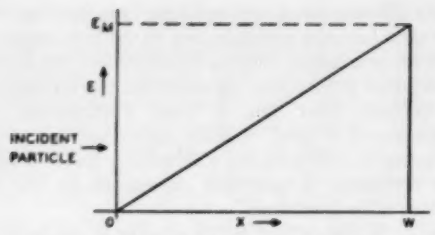


FIG. 1. Field distribution in np step junctions.

which cause the avalanche come from that direction in both an isolated reverse biased junction or in the collector of a transistor made by the alloy diffusion method which contains such a junction. Equation (4) then becomes

$$1-\frac{1}{M}=\tfrac{1}{2}W_1^2\int_0^{E_M}\alpha_i\exp\left[-\tfrac{1}{2}W_1^2\int_0^{E}(\alpha_i-\beta_i)dE'\right]dE. \tag{7}$$

The quantity

$$\alpha_i\exp\left[-\tfrac{1}{2}W_1^2\int_0^{E}(\alpha_i-\beta_i)dE\right],$$

which appears in the above equation can be thought of as an effective ionization rate. This ionization rate is no longer only a function of E. It is a function of the width constant of the junction and the electric field. For the same W_1 and V/V_B values, M would be different depending on whether the avalanche started with holes or electrons since the effective ionization rate expression is not symmetrical in α_i and β_i.

Equation (7) is differentiated with respect to E_M, yielding:

$$\alpha_i(E_M)=\frac{2}{W_1^2}\frac{d(1-1/M)}{dE_M}\exp\left[\tfrac{1}{2}W_1^2\int_0^{E_M}(\alpha_i-\beta_i)dE\right]. \tag{8}$$

The expression $(2/W_1^2)[d(1-1/M)/dE_M]$ reduces to $(2n/W_B)(E_M/E_{MB})^{2n-1}$ when the empirical analytic form for M is used, and Eq. (8) becomes

$$\alpha_i(E_M)=\frac{2n}{W_B}\left(\frac{E_M}{E_{MB}}\right)^{2n-1}\exp\left[\tfrac{1}{2}W_1^2\int_0^{E_M}(\alpha_i-\beta_i)dE\right], \tag{9}$$

where W_B is the width of the junction and E_{MB} the maximum field in the junction at breakdown.

Equation (9) was derived for the case in which the initial particles entering the junction are electrons. The analogous expression for holes is obtained by interchanging the α's and β's throughout.

$$\beta_i(E_M)=\frac{2n}{W_B}\left(\frac{E_M}{E_{MB}}\right)^{2n-1}\exp\left[\tfrac{1}{2}W_1^2\int_0^{E_M}(\beta_i-\alpha_i)dE\right]. \tag{10}$$

In these expressions, the n value and the breakdown voltages (and hence W_B and E_{MB}) appropriate to different junctions can be determined by experiment. The exponential term, however, cannot be evaluated directly. It can be determined from the solution of a differential equation derived from the experimental data.[7]

Equations (9) and (10) are a set of simultaneous integral equations which in principle determine the values of α_i and β_i. The problem of evaluating the exponential term can be simplified considerably by choosing two junctions with the same net density of impurity centers on the high-resistivity side, i.e., with the same W_1. Two such junctions, a pn^+ and an np^+, will be called complementary junctions. The n values appropriate to these junctions will be designated respectively by n_α and n_β. There is an additional benefit from this choice since E_{MB} has been found to be the same within experimental accuracy for complementary junctions.

It is shown in the Appendix in a purely mathematical exercise that a differential equation,

$$H''-\left(\frac{n_\alpha}{n_\beta}\right)y^{(n_\alpha/n_\beta)-1}H=0, \tag{11}$$

can be derived from a set of such integral equations. Here

$$y=(E_M/E_{MB})^{2n_\beta},$$

and

$$-\frac{H'(y)}{H(y)}=\exp\left(-\tfrac{1}{2}W_1^2\int_0^{E_M}(\alpha_i-\beta_i)dE\right).$$

The solution of this equation, subject to the appropriate boundary condition, gives the exponential term in Eqs. (9) and (10) in the form of Bessel functions which can be evaluated by series expansion. Therefore a complete solution for $\alpha_i(E_M)$ and $\beta_i(E_M)$ for any value of $E_M\leq E_{MB}$ is obtainable from the knowledge of n_α, n_β and the breakdown voltage for a set of complementary junctions.

EXPERIMENT

A series of experiments has been performed with transistors to determine n in the expression for multiplication (3). The transistors were made by the alloy diffusion method[8] which gives step junctions similar to those discussed above. The carriers were injected into the reverse biased collector junction from the nearby emitter junction. The current arriving at the collector is given by the emitter junction current times the current transport efficiency or current gain of the transistor, α. The multiplication measurements could then be made by determining the charge flowing across

[7] The author is indebted to P. A. Wolff for this method of attack.

[8] J. S. Saby and W. C. Dunlap, Jr., Phys. Rev. **90**, 630 (1953).

the collector barrier as a function of the bias with the emitter current held constant. Figures 2 and 3 give the results of such measurements on collector characteristics and the values of n determined in this way for representative pnp and npn transistors. For reasons cited below, this is at best an approximate way to determine the parameter n appropriate to a given junction. However the difference between the multiplication of the two collector junctions shown is unmistakable. These values are inaccurate because a slight uncertainty in the body breakdown voltage, V_B, would affect n greatly. Furthermore, the alpha of a transistor varies with the bias on the collector junction for several reasons. The principal effect is that caused by changes in the majority carrier flow pattern in the base with increasing multiplication of the collector. There are voltage drops in the base region which tend to concentrate the emission of minority carriers at the center of

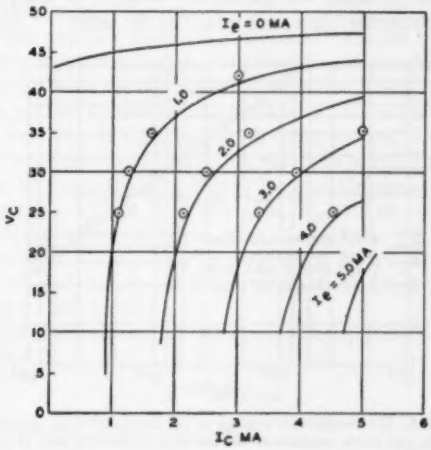


FIG. 2. Collector characteristics for a pnp transistor with about 0.7 Ω-cm base layer resistivity. ⊙ designates points calculated from $M=[1-(V/V_B)^{2.6}]^{-1}$.

the emitter and also as a secondary effect, give rise to a sweeping field toward the collector. This causes alpha to rise with increasing multiplication. In addition to this, the increase in the width of the collector depletion region with voltage brings the collecting boundary closer to the emitter with a consequent increase in alpha. This latter effect can be minimized in transistors with wide base layers.

A far more efficient and exact method for determining n is given below. The breakdown voltage is measured between the base and collector and between the emitter and collector (with the base floating) on a group of transistors made with the same resistivity base material. These two quantities are different. In the former case the breakdown voltage of the collector junction, that voltage at which $M=\infty$, is seen. In the latter case an apparent breakdown is observed at the voltage V_M,

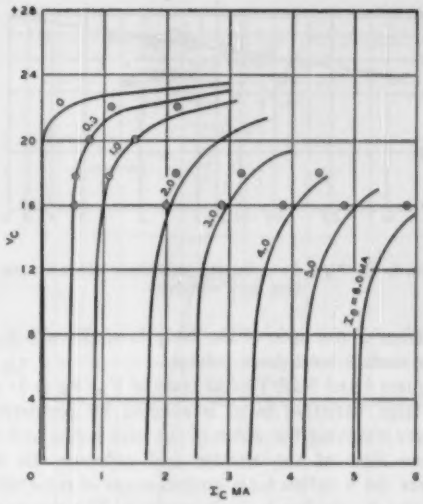


FIG. 3. Collector characteristics for an npn transistor with about 0.25 Ω-cm base layer resistivity. ⊙ designates points calculated from $M=[1-(V/V_B)^4]^{-1}$.

at which the low voltage α of the transistor, α_0, times the multiplication in the collector junction equals one, that is

$$\alpha_0 M(V_M)=1. \tag{12}$$

From Eqs. (12) and (3), the relation

$$1-\alpha_0=(V_M/V_B)^n \tag{13}$$

is obtained. Thus if $\log(V_M/V_B)$ is plotted *vs* $\log(1-\alpha_0)$ for each group of transistors made on the same material, the points should fall on a straight line of slope $1/n$. The experimental data when plotted in this way do define straight lines. Furthermore, the straight lines go through the $V_M/V_B=1$, $1-\alpha_0=1$ point as they should if the form of equation (3) is correct. Should the breakdown voltage of the junctions actually have been consistently due to a surface breakdown which is lower than the body breakdown of the junction this method of treating the data would have automatically detected it. The line through the experimental points would then have been parallel to the true line but above it by the

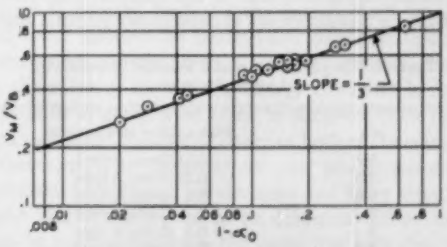


FIG. 4. V_M/V_B *vs* $1-\alpha_0$ for pnp transistors with 0.1 Ω-cm base layer resistivity.

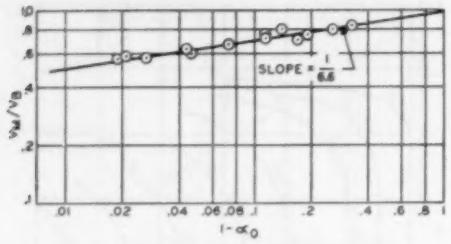


FIG. 5. V_M/V_B vs $1-\alpha_0$ for npn transistors with 0.5 Ω-cm base layer resistivity.

logarithm of the ratio of the body breakdown voltage to the surface breakdown voltage.

Figures 4 and 5 are typical plots of V_M/V_B vs $1-\alpha_0$. The large variation in α_0 is effected by geometrical changes involving the width of the base region and the relative sizes of the emitter and collector. In this manner the n values for a limited range of resistivities of n and p type has been determined. The ranges are restricted by the limitations of the present transistor technology. Table I gives the n values appropriate to step junctions vs the resistivity and resistivity type of the high-resistivity side. The npn transistors were made by alloying Pb As buttons onto wafers of p-type Ge, while the pnp transistors were made by the indium-alloying process on n-type wafers.

It should be noted that, in these measurements of the functional form of the multiplication, purely transistor effects have been minimized. All of the measurements are taken at the point where the alpha of the transistor times the multiplication equals unity or the base current is zero. Therefore there are no transverse majority carrier base currents to change the geometrical distribution of emission from the emitter junction.

The width of the space-charge region and the maximum field in the junction at breakdown are determined from the breakdown voltage as a function of the width constant. Step junctions of both the alloy type and grown-single-crystal type have been investigated for the breakdown point. Figure 6 gives the result vs the net number of impurity centers on the high-resistance side of the junction. The number of impurity centers on the high-resistance side of the junction was determined from measurements of capacitance per unit area.

TABLE I. The parameter n as a function of resistivity and resistivity type.

n	High ρ side of step junction
4.7	0.15 Ω-cm, p type
5.5	0.25 Ω-cm, p type
6.6	0.5 Ω-cm, p type
~6	2 Ω-cm, p type
3	0.1 Ω-cm, n type
3.4	0.6 Ω-cm, n type
3	2 Ω-cm, n type

All the junctions, whether grown or alloy or having n or p type on the high-resistivity side, fall on the same line on this log log plot. It is not incompatible with a difference between the ionization rates of holes and electrons that np^+ and pn^+ junctions fall on the same curve. At very high multiplications the number of electrons and holes participating in the avalanche are essentially equal and so the average effective ionization rate would be very close to the average between that for electrons and holes individually. In a junction across which the field is constant as a function of distance, it can be shown rigorously that the breakdown voltage would be the same regardless of whether the initiating particles are electrons or holes. Since the ionization rates are a very rapidly rising function of field strength, the high-field end of a step junction can be approximated by a plateau of constant field and the rest of the junction neglected. Insofar as this approximation is valid, the breakdown voltages of a set of complementary junctions would be the same.

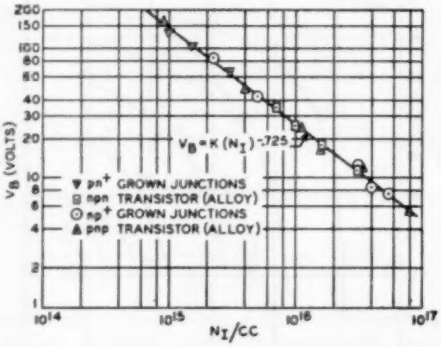


FIG. 6. The breakdown voltage vs the net number of impurity centers per cubic centimeter on the high resistivity side of step junctions.

The equivalence in breakdown properties between alloy and grown junction indicates that whatever strains or imperfections the alloying process introduces near the junction ordinarily do not affect the body breakdown process. The slope of the breakdown voltage plot is approximately -0.725 whereas it should have been -1.0 if the Zener theory of breakdown held. Since there is no tendency toward this latter slope even at the highest impurity concentrations, which correspond to the highest electric fields, it is considered that this plot gives strong evidence that the internal field emission has not been observed in this experiment. The lowest breakdown voltage observed corresponds to a maximum electric field of 320 000 volts/cm. Therefore internal field emission does not occur below this field strength in germanium.

A serious experimental problem in this type of investigation is the differentiation between breakdown on the surface of the semiconductor and in the body of

the junction. The measurement of the multiplication below breakdown is one method which has already been mentioned. However, there is considerable experimental evidence that many surface breakdowns are also multiplicative. But, in experiments like those discussed above for the determination of the parameter n, very few of the carriers coming from the emitter would be collected by such a surface breakdown region. In the case of isolated junctions, measurements of the resistance after breakdown is a good indication. Breakdowns occurring near the surface in regions of small cross section have considerably higher series resistance than body breakdowns. Therefore pulse measurements of voltage at high currents can be used to separate the two. For ordinary junctions there will always be a small positive resistance in the breakdown region because of the body resistance of the semiconductor on both sides of the junction and because of the effect of the space charge of the carriers in the high field region with increasing current.

The quantity $2n/W_B$ is determined from the breakdown and multiplication data for different junctions in the limited range of resistivities accessible. These values are plotted against the maximum field in the junction at breakdown, E_{MB}, in Fig. 7. From curves through these experimental points, the values of n and $2n/W_B$ can be estimated very closely for two complementary junctions. (These have the same E_{MB}.) Two sets of complementary junctions were chosen. Set 1 has a maximum electric field at breakdown of 237 kv/cm, $n_\alpha \simeq 6$ and $n_\beta \simeq 3$. Set 2 has an E_{MB} of 260 kv/cm, $n_\alpha \simeq 4.5$ and $n_\beta \simeq 3$.

For the first set of junctions, the equation to be solved is

$$H''-2yH=0.$$

The solution of this equation is

$$\exp\left(-\tfrac{1}{2}W_1^2\int_0^{E_M}[\alpha_i(E)-\beta_i(E)]dE\right)=-\frac{H'(y)}{H(y)}$$
$$=\frac{1-y+\frac{2}{3}y^3-(2/15)y^5+(1/18)y^6\cdots}{1-y+\tfrac{1}{3}y^3-\tfrac{1}{6}y^4+(1/45)y^6-(1/126)y^7\cdots}.$$

For the second set of junctions the solution of the appropriate equation is

$$\exp\left(-\tfrac{1}{2}W_1^2\int_0^{E_M}[\alpha_i(E)-\beta_i(E)]dE\right)=-\frac{H'(y)}{H(y)}$$
$$=\frac{1-y^{3/2}+\frac{2}{5}y^{5/2}-(15/100)y^4+(9/175)y^5\cdots}{1-y+\frac{2}{5}y^{5/2}-(6/35)y^{7/2}+(3/100)y^5-(3/350)y^7}.$$

With the use of these functions, the values of $\alpha_i(E_M)$ and $\beta_i(E_M)$ were determined for (E_M/E_{MB}) equal to 0.95, 0.9, 0.8, and 0.7 in both cases. The series are highly convergent for these values since $y=(E_M/E_{MB})^{2n\beta}$. The

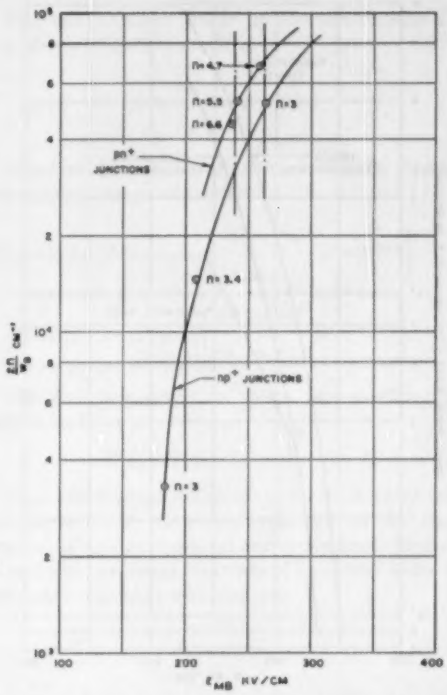


FIG. 7. The quantity $2n/W_B$ *vs* the maximum field in the step junction at breakdown. The dashed lines represent the two sets of complementary junctions chosen for the ionization rate calculations.

results of these calculations are plotted in Fig. 8. The electron ionization rate curve is about a factor of two below the curve for holes in this region. Points obtained from the two sets of junctions are in excellent agreement with each other and are therefore consistent with the assumption that the ionization rate per unit path length is primarily only a function of field strength.

A theoretical curve calculated by Wolff for an assumed mean free path between electron (or hole)—phonon collisions of 130A and a threshold for electron-hole pair production of 1.5 ev is also given in Fig. 8. This curve corresponds almost exactly with the experimental ionization rate curve for holes while the electron curve is parallel and slightly below. Such a small difference between the electron and hole ionization rate corresponds to only about a 5 to 10% positive change in the value of the ratio of energy threshold to mean free path in going from holes to electrons. This could be made up of changes in either or both of the relevant parameters. The threshold for pair production could certainly be different for electrons and holes since the band shapes are known to be different.[9] There could also be some variation in the mean free path due to a

[9] C. Kittel, Physica **20**, 829 (1954).

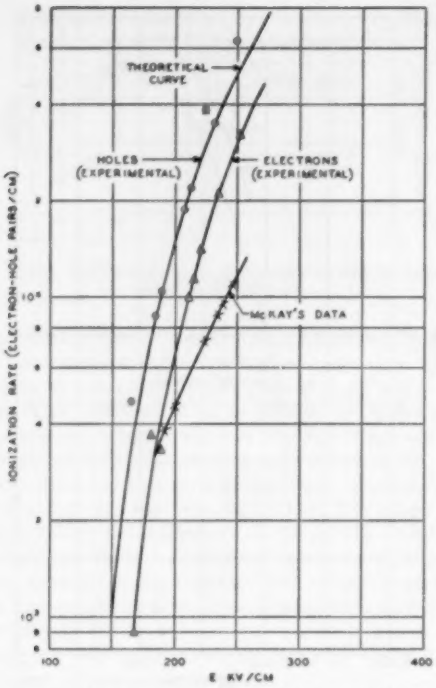


FIG. 8. Ionization rate *vs* electric field strength for Ge. ⊙ calculated from the experiment for holes. △ calculated from the experiment for electrons. × obtained by McKay from his experiment on a linear graded Ge junction. The experimental curve for holes was drawn to coincide with the theoretical curve.

difference in the coupling of electrons and holes to the optical phonon field.

The values of 1.5 ev and 130 A for energy threshold and mean free path are to be compared with the values 2.3 ev and 130A obtained for silicon by Wolff[9,10] from McKay's data. These sets of parameters are consistent with each other in that in both cases the energy threshold is slightly above twice the energy gap.

It would be appropriate here to examine the physical reason why such a relatively small difference between the ionization rates for holes and electrons is reflected in such a large difference in the empirical multiplication laws for np^+ and pn^+ junctions. In a pn^+ junction where the initial current is made up of electrons, the average particle ionization rate starts on the electron curve for low voltages or electric fields (and low multiplication) and as the voltage and multiplication increase the average particle ionization rate rises toward the hole curve since holes make up a higher and higher proportion of the particles participating in the avalanche. Therefore the multiplication rises very sharply with increasing voltage. On the other hand, similar arguments would tend to make the rise in multiplication of np^+ junctions far more gradual.

In Fig. 8 are also plotted some ionization rates *vs* electric field obtained by McKay on a linear graded germainium junction by his method. The agreement between the two methods is rather good. The discrepancy is probably due to the neglect of the difference in ionization rates for electrons and holes and the greater experimental difficulties and uncertainties attendant upon McKay's experiment.

CONCLUSIONS

Data on breakdown and multiplication *vs* bias of germanium junctions have been used to obtain ionization rates for both holes and electrons in high fields. The ionization rate for holes is greater than that for electrons by about a factor of two. The variation of ionization rate with field strength is in excellent agreement with the theoretical predictions of Wolff.

A by-product of this agreement with theory is a determination of the mean free path for electron (or hole)—phonon collisions and the energy threshold for pair production in Ge. These are respectively about 130A and 1.5 ev. These numbers individually are subject to considerable uncertainty but their ratio is probably quite good.

The experimental evidence is strong that there is no field emission in germanium up to a field strength of at least 320 000 volts/cm. Therefore it is concluded that all breakdowns (not occurring on the surface) of germanium junctions above about 5 volts are the result of McKay's avalanche process, although the multiplication inherent in such a process was actually measured in only some of the junctions. These did, however, cover the whole range investigated.

ACKNOWLEDGMENTS

This work has been made possible by the cooperation of a great many members of the Bell Telephone Laboratories Staff. Particular acknowledgment is due W. C. Meyer who gathered most of the experimental data and J. J. Ebers, J. L. Moll, J. M. Early, and P. A. Wolff for many enlightening discussions on this subject. Thanks are also due K. G. McKay for making his unpublished data on germanium available and for a critical reading of the final manuscript.

APPENDIX

Equations (9) and (10) of the text are rewritten using the relation $W=\frac{1}{2}W_i^2E_M$ which can be derived from (1). Then

$$\frac{2n_a}{E_{MB}}\left(\frac{E_M}{E_{MB}}\right)^{2n_a-1}=\frac{W_i^2}{2}\alpha_i(E_M)I(E_M), \qquad (14)$$

[10] P. A. Wolff (personal communication). The value for the mean free path has been corrected from that appearing in reference 5.

and

$$\frac{2n_\beta}{E_{MB}}\left(\frac{E_M}{E_{MB}}\right)^{2n_\beta-1}=\frac{W_1^2}{2}\frac{\beta_i(E_M)}{I(E_M)}, \tag{15}$$

where

$$I(E)=\exp\left(-\frac{W_1^2}{2}\int_0^E(\alpha_i-\beta_i)dE\right).$$

But

$$\frac{W_1^2}{2}\alpha_i(E_M)I(E_M)=-\frac{dI(E_M)}{dE_M}+\frac{W_1^2}{2}\beta_i(E_M)I(E_M).$$

When this relation is substituted into Eq. (14) it becomes

$$\frac{dI(E_M)}{dE_M}-\frac{2n_\beta}{E_{MB}}\left(\frac{E_M}{E_{MB}}\right)^{2n_\beta-1}[I(E_M)]^2=-\frac{2n_\alpha}{E_{MB}}\left(\frac{E_M}{E_{MB}}\right)^{2n_\alpha-1}. \tag{16}$$

Now a new variable $y=(E_M/E_{MB})^{2n_\beta}$ is introduced. Equation (16) then simplifies to

$$\frac{dI(y)}{dy}-[I(y)]^2=-\frac{n_\alpha}{n_\beta}y^{(n_\alpha/n_\beta)-1}. \tag{17}$$

This can be made linear by the substitution $I(y)=-H'(y)/H(y)$. The result is

$$H''-\frac{n_\alpha}{n_\beta}y^{(n_\alpha/n_\beta)-1}H=0. \tag{18}$$

Equation (18) is a variation of Bessel's equation whose solutions are of the form

$$H(y)=K_1(\sqrt{y})J_{n_\beta/(n_\alpha+n_\beta)}\left[\frac{2i(n_\alpha/n_\beta)^{\frac{1}{2}}}{(n_\alpha/n_\beta)+1}y^{\frac{1}{2}(n_\alpha/n_\beta)+\frac{1}{2}}\right]$$

$$+K_2(\sqrt{y})J_{-n_\beta/(n_\alpha+n_\beta)}\left[\frac{2i(n_\alpha/n_\beta)^{\frac{1}{2}}}{(n_\alpha/n_\beta)+1}y^{\frac{1}{2}(n_\alpha/n_\beta)+\frac{1}{2}}\right]. \tag{19}$$

The boundary condition which this solution must obey is that

$$-H'(y)/H(y)=1 \quad \text{when} \quad y=0.$$

Since the highest value of y which is of physical interest is unity, the series expansion for the Bessel function is highly convergent and can be used effectively to evaluate the Bessel functions of fractional order and imaginary argument which appear.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Gyromagnetic Ratio of Iron at Low Magnetic Intensities

G. G. Scott
Research Laboratories Division, General Motors Corporation, Detroit, Michigan
(Received Februrary 24, 1955)

It is shown that the measured value for the gyromagnetic ratio of pure iron as determined by a direct magneto-mechanical method, undergoes a change for low values of the induced magnetic intensity. The effective value of this ratio extrapolated to zero intensity, checks the value which is theoretically expected from a consideration of recent ferromagnetic resonance experiments. For higher values of the induced magnetic intensity, the gyromagnetic ratio approaches a constant value which checks the value obtained in previous investigations based upon the Einstein-de Haas effect.

INTRODUCTION

FOR a number of years an investigation of the gyromagnetic ratios of the iron-nickel alloy series has been under way at the General Motors Research Laboratories. It was noted during this investigation that the measured value of the gyromagnetic ratio (ρ) for a given alloy always increased for low values of magnetic intensity. Although the shape of the curve of magnetic intensity *vs* gyromagnetic ratio appeared to be different for different concentrations of nickel in iron, there was nevertheless, always an increase in ρ for small induced magnetic intensities. Also when the largest obtained ρ values were plotted against concentration of nickel in iron it was found that all points from 0 percent Fe 100 percent Ni to 90 percent Fe 10 percent Ni fell on a smooth curve. The value for pure iron which had been previously determined was, however, an exception. It was, therefore, decided to undertake an extensive investigation of pure iron varying the induced magnetic intensity, and going down to as low a value of magnetic intensity as was practicable. It was found that a similar effect exists for pure iron also. Furthermore, when the curve of ρ *vs* intensity, was extrapolated to zero intensity, the gyromagnetic ratio was found to have a value which checks the value which is theoretically expected from a consideration of recent work in ferromagnetic resonance. For high values of magnetic intensity the gyromagnetic ratio approaches a constant value, which checks previous work done at the General Motors Research Laboratories, and also the work of other investigators.

TABLE I. Condensed data for determination of gyromagnetic ratios (ρ).* $\rho\frac{e}{m}=\left(\frac{\pi Id}{4Pzk(m/e)}-2i_s\Sigma A_s\right)\Big/(M_s-i_s\Sigma A_s)$.

i_s	M_s	P	d	$\rho e/m$
4.0004	20832	27.517	0.034141	1.0465
2.0005	9695	27.520	0.016021	1.0542
2.0001	9693	27.515	0.016068	1.0579
10.0031	54571	27.532	0.089126	1.0429
7.0008	37676	27.536	0.061691	1.0453
7.0006	37674	27.536	0.061646	1.0446
1.0000	4276	27.520	0.007099	1.0572
1.0000	4276	27.520	0.007049	1.0495
4.0005	20832	27.523	0.034198	1.0481
2.0002	9694	27.515	0.016126	1.0618
2.0005	9695	27.515	0.016085	1.0588
1.0000	4276	27.514	0.007150	1.0653
1.0000	4276	27.515	0.007125	1.0614
1.0000	4276	27.515	0.007145	1.0644
10.0033	54572	27.532	0.088769	1.0386
3.0002	15213	27.487	0.025025	1.0509
3.0000	15212	27.490	0.025057	1.0522
3.0000	15212	27.490	0.025003	1.0499
5.0007	26415	27.486	0.043237	1.0462
5.0005	26414	27.486	0.043359	1.0492
6.0004	32017	27.493	0.052351	1.0449
2.5001	12412	27.482	0.020482	1.0542
2.5000	12411	27.481	0.020476	1.0535
2.5002	12412	27.481	0.020563	1.0585
3.5003	17980	27.481	0.029491	1.0482
3.5003	17980	27.481	0.029597	1.0521
2.5000	12411	27.483	0.020587	1.0598
3.5001	17979	27.485	0.029548	1.0502
5.0001	26412	27.484	0.043331	1.0485
4.0003	20802	27.485	0.034135	1.0487
4.0001	20801	27.481	0.034061	1.0466
5.0003	26413	27.482	0.043230	1.0462
5.0003	26413	27.482	0.043287	1.0476
7.0016	37628	27.493	0.061746	1.0489
7.0010	37624	27.493	0.061557	1.0457
1.0001	4249	27.477	0.007074	1.0614
1.0003	4250	27.477	0.007148	1.0726
1.0000	4249	27.484	0.007161	1.0749
1.0001	4249	27.484	0.007093	1.0641

Moment of inertia, first group	$I=203.16$ g cm^2
Moment of inertia, second group	$I=203.08$ g cm^2
Optical length	$z=1576.9$ cm
Phase angle constant	$k=0.99941$
Winding constant	$\Sigma A_s=78115$ cm^2
Mass-charge ratio of electron	$m/e=5.6844\times10^{-9}$ g coulomb^{-1}

* i_s = magnetizing current, milliamperes. M_s = magnetic moment, amp cm^2. P = period, sec. d = amplitude change per reversal of i_s, cm.

EXPERIMENTAL PROCEDURE

The general experimental procedure used in these experiments has been previously described.[1] However, since there have been numerous improvements in the technique it appears desirable to give a brief outline of the method used to obtain the major experimental factors required to determine the gyromagnetic ratios in this experimental work.

Angular Momentum Changes

In Table I each value of gyromagnetic ratio given is the result of taking eight different sets of data on change of angular momentum. In four of these sets the amplitude of the torsional pendulum is built up by resonance using square wave impulses, and in four more the amplitude is similarly cut down.

The period of the torsional pendulum is about 27 seconds and the square wave is obtained by using a hand operated reversing switch. Before starting a series of runs a current considerably above any that is ever used in the actual experiment is passed through the magnetizing winding and gradually reduced in strength to zero while the reversing switch is rapidly reversed. This demagnetization process is necessary in order to obtain a high degree of correlation between the change in magnetizing current and the resulting change in magnetic moment.

The amplitude of the swinging torsional pendulum is built up for 30 complete periods; it is then cut down for 30 complete periods. All of the amplitude values are recorded for both sets. The azimuth is then changed 180° and two more sets obtained in a similar manner. The second half of the data required for one value is obtained by reversing the switch located at the lower end of the element[1] again demagnetizing, and repeating the aforementioned steps.

The data given in the tables were taken in two groups. The first group was obtained between November 11, 1954 and November 26, 1954. The second group was obtained between December 13, 1954 and January 7, 1955.

Data correlating changes in magnetizing current with changes in magnetic moment were taken both before and after each one of these two groups.

The moment of inertia was measured at the end of each group.

Magnetic Moment Determinations

The changes in magnetic moment occurring in these experiments are correlated with changes in the current flowing in the magnetizing winding which is wound on the rod. To do this a separate experiment is performed wherein the wound rod is supported in a horizontal position on a critically damped torsional system.

The apparatus is arranged so that the gyromagnetic rod can be readily replaced by an air core coil whose dimensions and number of turns are accurately known. A current of 16 milliamps is passed through the air core coil; the resulting magnetic moment can then readily be computed. This is the standard of magnetic moment which has been used for all of the work on gyromagnetic ratios at this laboratory.

In determining magnetic moments for the gyromagnetic rod, the standard is first placed in the torsional system. The current in the large horizontal NS Helmholtz coil system is then adjusted so that a torsional deflection of 50.00 cm is obtained on the 52 ft optical lever system, when the current in the air core standard is reversed. The air core standard is then replaced by the gyromagnetic rod. This rod is first demagnetized, and then the desired current is passed through its

[1] G. G. Scott, Phys. Rev. 82, 542 (1951).

magnetizing winding. After several preliminary reversals, the current in the NS system is again adjusted so as to obtain 50.00 cm upon reversal of the current in the gyromagnetic winding. Allowance is, of course, made for the earth's horizontal field. This is done by taking these readings both with and against the direction of the earth's horizontal component. The magnetic moment of the rod is found by multiplying the magnetic moment of the standard by the ratio of the NS field values required to produce the common torque. A separate determination of magnetic moment is made for each current value that is used in the experiments to determine changes in angular momentum.

Moment of Inertia

In determining, moment of inertia, a series of four brass cylinders is used. These cylinders are chosen to differ in mass by a few grams. The mass and radius of each cylinder is measured and its moment of inertia computed. The period of at least 3 of these cylinders and that of the unknown gyromagnetic pendulum system is measured on the same suspension that is used in the angular momentum experiment. The mass of the gyromagnetic pendulum system is also carefully determined. This mass lies within a few grams of that of the cylinders. From the periods of the three known cylinders the variation in torsional constant of the suspension ribbon with suspended mass can be determined. This can then be applied to the unknown pendulum system to determine the desired moment of inertia. Accuracy of 1 part in 2000 is readily obtainable.

Period

The period of the swinging pendulum system is determined once each day. Time signals from WWV are utilized, and an elapsed time between readings of from ½ hour to 1 hour is used. High accuracy is easily obtainable.

Element Current

The current which is reversed in the magnetizing winding on the gyromagnetic rod is measured to five places on an L and N type K potentiometer. This is done both before and after each "build up" and "cut down" run. This gives a total of 16 current readings for each value of gyromagnetic ratio reported in the table.

INSTRUMENTATION

The instrumentation used in this experiment is much the same as has been recently reported (reference 1). That which has been changed will be briefly described.

Vacuum Chamber

In previous work on measuring gyromagnetic ratios it was not possible to support the ferromagnetic rod in a horizontal position in the vacuum chamber which was used for the measurement of angular momentum changes. It was, therefore, necessary to use a separate chamber for the measurement of magnetic moment. This situation has been remedied by the construction of a new vacuum chamber in which the central section is somewhat spherical in shape, so that the rod may be supported in a horizontal as well as in a vertical position. Other changes were made so as to increase the accessibility of the pendulum system for the purpose of making the necessary adjustments.

Gradient Detector

A gradient detector was constructed. A closely matched and accurately aligned pair of coils was made into an astatic system and supported on a critically damped torsional system in the main vacuum chamber. The current through this coil system was reversed and the resulting rotation observed on the regular 52-ft optical system. The sensitivity of this system was such that a vertical gradient in the horizontal component of 8×10^{-8} gauss per cm, produced a 1-mm deflection of the system. The normal uniform horizontal component of the earth's field produced only a few mm deflection.

By the use of this instrument it was found that several pieces of ferromagnetic material used in the building construction which were previously considered to be far enough away from the main instrument, were causing gradients several times larger than the gradients inherent to the large Helmholtz coil system. This ferromagnetic material was, of course, removed.

Helmholtz Coil System

The homogeneity of the magnetic fields produced by the 8 ft and 9 ft diameter Helmholtz coil systems has been considerably improved by using in connection with each one of these systems an additional Helmholtz coil system concentric with the large system but about ½ the size. The field produced by the small coil set is made to oppose that produced by the large system. By an inspection of the equations for the fields produced by a Helmholtz coil pair[2] it can readily be seen that two systems of different sizes connected in series opposition can be made to produce a much more uniform field than that produced by the large one alone, if the number of turns on the small and large systems is proportional to the fifth power of the radii of the two systems.

That a very considerable increase in homogeneity was achieved, was verified by measurements made with the gradient detector described previously.

RESULTS

In Table I is given a condensation of all of the data taken in making these gyromagnetic determinations. Table II gives an average value for $\rho e/m$ for each

[2] Ruark and Peters, J. Opt. Soc. Am. 13, 205 (1926).

TABLE II. Gyromagnetic ratios (ρ) and Landé factors (g') for various induced magnetic intensities in iron.[a]

i_c	β	$\rho e/m$	g'
1.0	107.7	1.063	1.881
2.0	245.3	1.058	1.890
2.5	314.2	1.056	1.893
3.0	385.2	1.051	1.903
3.5	455.4	1.050	1.904
4.0	527.4	1.048	1.909
5.0	669.3	1.047	1.909
6.0	811.4	1.045	1.914
7.0	954.3	1.046	1.912
10.0	1383.5	1.041	1.922

[a] i_c = winding current, milliamperes. β = average induced magnetic intensity, amp cm^{-1}. $\beta = (M_s - i_c \Sigma A_c)/v$, $v = 38.88$ cm^3.

value of the magnetizing current used. The corresponding values of the Landé factor (g') are also given in Table II. Figure 1 is a plot of the data shown in Table II. In this curve g' is plotted against the average induced magnetic intensity of the rod. The average magnetic intensity is obtained by dividing the total moment of the rod in amp cm^2 by the volume of the rod.

Since this rod is magnetized by a uniformly distributed winding the resulting induced magnetization will, of course, not be uniform but will rather approximate an elliptical distribution of magnetic intensity. This nonuniformity of magnetization, of course, has its effect on the shape of the g' *vs* intensity curve. It is planned to repeat this experiment using a magnetizing winding of a shape that will more closely induce uniform intensity of magnetization. Although this should change the shape of the curve, it is considered unlikely that either the extrapolated zero intensity value or the limiting value at higher intensities, would be effected by such a change in winding shape.

The value of g' extrapolated to zero intensity is 1.873 ± 0.005. This value is in good agreement with the value expected from recent experiments in ferromagnetic resonance.[3]

The limiting value at higher values of magnetic intensity appears to be about 1.919, corresponding to $\rho e/m$ of 1.042 ± 0.003. This is in reasonable agreement with previous work at this laboratory and with the work of Barnett and Meyer.[4] It should be here noted that the values given in reference 4 differ slightly from those previously reported by the present author. It was pointed out by Dr. Sheldon Brown that a small systematic error involving the optical system existed in the original work. Corrected values are given in reference 4.

CONCLUSION

It has been shown that the effective value of g' as measured by an Einstein-de Haas experiment, changes from one value at zero intensity of magnetization to a higher value as the magnetic intensity increases. The value obtained by extrapolation to zero magnetic intensity checks the value to be expected by a consideration of recent work in the field of ferromagnetic resonance. For higher magnetic intensities the effective value of g' approaches a limiting value which checks previous work done at the General Motors Laboratory and also the work of other investigators who have determined gyromagnetic ratios by magneto-mechanical methods. The zero intensity value for g' from the present series of experiments is 1.873 ± 0.005.

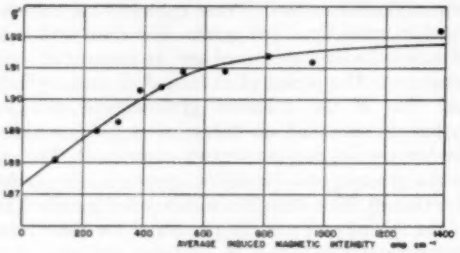


FIG. 1. Plot of Landé Factor (g') *vs* average induced magnetic intensity. Iron rod 1.5-cm diameter, 22 cm long.

This same effect has been noted for seven different alloys in the iron nickel series. In each case a larger g' was noted for larger values of the intensity of magnetization. This work on alloys will be reported when more information has been obtained.

The author wishes to express his appreciation to Dr. E. J. Martin and to the other members of the staff of the General Motors Research Laboratories, whose cooperation made this work possible. Special thanks are also due Mr. Robert Frank for the large amount of help given in obtaining the experimental data, and in making the subsequent calculations.

[3] C. Kittel, *Introduction to Solid State Physics* (John Wiley and Sons, Inc., New York, p. 171.

[4] Brown, Meyer, and Scott, Compt. rend. **238**, 2502 (1954).

Electrical Properties of Plastically Deformed Germanium

A. G. TWEET
General Electric Research Laboratory, Schenectady, New York
(Received May 6, 1955)

Samples of Sb-doped (10 ohm-cm *n*-type) and Au-doped (20 ohm-cm *p*-type) Ge have been plastically deformed at 550°C–620°C. The deformation created a large concentration of acceptor centers distributed inhomogeneously throughout the deformed sections of the samples. These acceptors lie near the valence band and converted the Sb-doped specimens to *p*-type. Hall coefficient measurements have been made as a function of temperature to determine the energies of these acceptors. Upon annealing at temperatures above 750°C from 30 minutes to 1½ hours, the acceptor levels near the valence band disappear, leaving only the acceptors probably associated with edge dislocations. It is postulated that the annealable acceptors may be due to the presence of vacancies.

I. INTRODUCTION

THERE has been considerable study recently of the electrical properties of plastically deformed Ge.[1–5] The experimental investigations have all indicated that plastic deformation at temperatures between 500°C and 700°C produces acceptor centers in the Ge, but there has been some disagreement as to whether or not the material can be converted to *p*-type by this treatment.

It has been indicated[4] that acceptor centers with energies less than 0.15 ev from the valence band are created during the plastic deformation of Au-doped Ge. The work reported here comprised a somewhat more extensive examination of this effect. It also included an investigation of certain annealing effects observed in the deformed material at elevated temperatures.

II. EXPERIMENTAL

A. Procedure

Bars of Sb-doped Ge (~10 ohm-cm, *n*-type) and Au-doped Ge (~20 ohm-cm, *p*-type) $\frac{7}{8}$ in.×$\frac{1}{8}$ in.×$\frac{1}{32}$ in. have been bent to a radius of curvature of $\frac{1}{4}$ in. on a quartz form in an atmosphere of 99.9 percent N_2 at temperatures between 550° and 620°C, as read on a Chromel-Alumel thermocouple. The bars were etched and cleaned with KCN to remove Cu before insertion in the bending apparatus, which was given the same KCN treatment. In order to simplify the experimental arrangements, the bars were also straightened in the same apparatus before proceeding with the electrical measurements. (Separate experiments indicated no qualitative difference between the electrical properties of the bent and straightened bars and those of the bars which were simply bent.)

Indium contacts were fused to the samples, and the Hall coefficient measured as a function of temperature in a permanent-magnet Hall effect apparatus.[6] Some Hall measurements were also made by using pressure contacts of phosphor bronze tinned with In. There was no appreciable difference between the results of measurements by the two methods, indicating that preferential diffusion of the In down the dislocations[7] very probably did not occur.

Other samples which had been bent and straightened were treated with KCN and annealed for various lengths of time at temperatures above 600°C in a 99.9 percent N_2 atmosphere. Hall coefficient measurements were then made on these annealed specimens as a function of temperature, by using fused tin contacts.

B. Results

Figure 1 shows the Hall coefficient, R, in cm³/coulomb plotted *vs* the reciprocal of the temperature for two sections of a sample of Au-doped Ge. Both sections were *p*-type, both before and after the deformation, with a room temperature hole concentration of ~10^{14}/cm³. The inset identifies the two sections, one of which has been plastically deformed. Note that the contacts are arranged for Hall effect measurements only. The straight line portion of Curve *A* is of the form $e^{+\varphi/kT}$, where φ=0.15 ev, characteristic of the lower Au acceptor state in Au-doped Ge.[8] In Curve *B* the straight line portion of the plot has a φ of 0.10 ev. These data are representative of data taken on several similar samples, which show a spread in φ from 0.08 ev to 0.12 ev.

Figure 2 is a Hall coefficient *vs* $1/T$ plot for a sample of plastically deformed Sb-doped Ge. Again two sections of the sample have been measured; both were in the deformed region in this case. Before deformation, the sample was uniformly 10 ohm-cm *n*-type, whereas after deformation both sections were *p*-type with a room temperature hole concentration of 2×10^{14}/cm³. These data show quite strikingly the extreme inhomogeneity which is usually observed in bent specimens. The slope of the straight line portion of Curve *A*, which is of the form $e^{+\varphi/kT}$, yields φ=0.10 ev. Again this is

[1] C. J. Gallagher, Phys. Rev. **88**, 721 (1952).
[2] W. C. Ellis and E. S. Greiner, Phys. Rev. **92**, 1061 (1953).
[3] Pearson, Read, and Morin, Phys. Rev. **93**, 666 (1954).
[4] C. J. Gallagher and A. G. Tweet, Phys. Rev. **96**, 834 (1954).
[5] W. T. Read, Phil. Mag. **45**, 775 (1954).
[6] A. G. Tweet, this issue [Phys. Rev. **99**, 1182 (1955)].
[7] D. Turnbull and R. E. Hoffman, Acta Metallurgica **2**, 419 (1954).
[8] W. C. Dunlap, Jr., Phys. Rev. **91**, 1282 (1953).

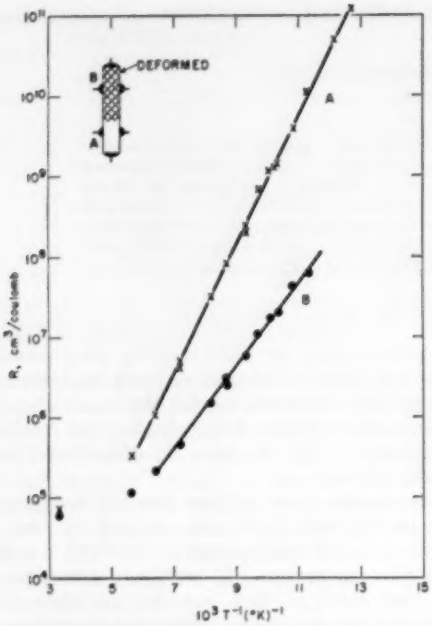


FIG. 1. Hall coefficient *vs* $1/T$ plots for plastically deformed (B) and undeformed (A) sections of a bar of Au-doped Ge. Both sections are p-type. Straight line portion of each curve is of the form $e^{\varphi/kT}$. $\varphi_A = 0.15$ ev, characteristic of lower Au acceptor level in Ge; $\varphi_B = 0.10$ ev, representative of acceptor levels introduced in Ge by plastic deformation.

characteristic of slopes observed on many such plots, which show a spread in φ from 0.08 ev to 0.12 ev. The data for Curve B are too meager to justify the assignment of a slope.

The reason for the relatively low resistance section shown in Curve B (Fig. 2) is not known. Such sections were found on roughly half of the samples investigated. In a few Au-doped samples, the entire bent region was of relatively low resistance, and it was possible to obtain data down to much lower temperatures. In these cases, the slopes of the log$R$ *vs* $1/T$ plot yielded a φ of $\sim$0.04 ev. (This is equal to the energy reported for the low-lying acceptor level due to Cu in Ge.[9] However, the unbent ends of these samples gave no evidence of Cu contamination.)

Figure 3 is a Hall coefficient *vs* $1/T$ plot for two specimens from one ingot of Sb-doped Ge which were bent, straightened and annealed. Two control specimens cut from adjacent sections of the same ingot were annealed with the deformed ones and are also plotted for comparison. The two deformed specimens, which converted from n- to p-type upon deformation, reverted to n-type during annealing. The control samples remained $\sim$10 ohm-cm n-type during annealing. Curve A is for a plastically deformed sample which was heated ½ hour at 800°C followed by 1½ hours at 600°C. Curve B is for a similar sample which was annealed at 750°C for 1 hour followed by 1 hour at 600°C. Curves A' and B' are for the control specimens which were given the same heat treatment as samples A and B, respectively. The dotted line is a plot of data of Pearson, Read, and Morin[3] on deformed n-type Ge.

Prior to annealing, the Hall coefficients of the deformed specimens were measured using pressure contacts and were found to be similar to those shown in Fig. 2 in their temperature dependence.

The annealing process reduced macroscopic sample inhomogeneity in the deformed samples very considerably, and it was possible to make fairly reliable Hall mobility measurements. In Fig. 4 are plotted mobility data on the four samples of Fig. 3, with the curves labeled as on Fig. 3. The mobility in the deformed and annealed specimens seems to be approximately proportional to the temperature, as predicted by Dexter and Seitz[9] on the basis of a deformation potential scattering mechanism. However, the magnitude of the scattering indicated by the experiments is very much larger than that predicted for the dislocation densities present in our samples.

Read[10] has calculated the effects of scattering by charged dislocations upon Hall mobility in semiconductors. This approach seems to be well-suited to dis-

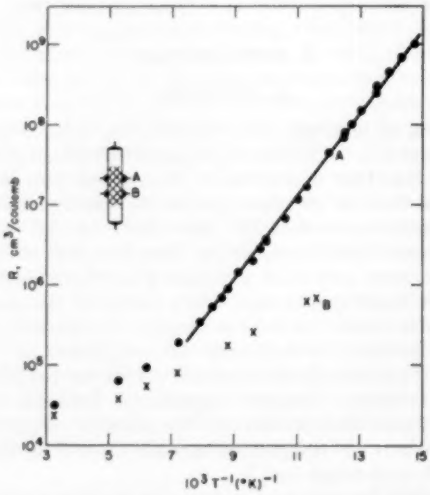


FIG. 2. Hall coefficient *vs* $1/T$ plot for two sections of a bar of plastically deformed Sb-doped Ge, 10 ohm-cm n-type before deformation, p-type after deformation. Straight line portion of Curve A is of the form $e^{\varphi/kT}$, $\varphi = 0.10$ ev. This behavior is representative of the effects of acceptor levels introduced in Ge by plastic deformation. Note the inhomogeneity, characteristic of most of the samples investigated.

[9] Fuller, Struthers, Ditzenberger, and Wolfstirn, Phys. Rev. **93**, 1182 (1954).

[9] D. L. Dexter and F. Seitz, Phys. Rev. **86**, 964 (1952).

[10] W. T. Read, Jr., Phil. Mag. **46**, 111 (1955).

location scattering below room temperature in the n-type Ge samples of Figs. 3 and 4.

Annealing effects have also been observed on deformed Au-doped Ge samples. In this case, annealing eliminated the behavior of the Hall coefficient in the deformed region shown in Curve B (Fig. 1), so that the carrier density now follows Curve A (Fig. 1) over the entire sample.

III. DISCUSSION

The experimental results described above are consistent with the idea that plastic deformation of Ge introduces, in addition to the array of edge dislocations required by the geometry of the deformation, some other crystalline defect which can be annealed out of the crystal at elevated temperatures. The shapes of the log$R$ *vs* $1/T$ curves for the deformed samples indicate that these additional defects introduce at least one kind of isolated localized acceptor level. The isolated nature of these acceptors must be contrasted with the strongly interacting acceptors postulated by Read[6] as being associated with edge dislocations in Ge. The latter, of course, yield an entirely different shape for the curve of R *vs* $1/T$.

It is believed that the annealing heat treatment of the deformed Ge, while removing the acceptors near the valence band, perhaps by removing the defects giving rise to them, does not result in an appreciable diminution in the density of dislocations in the material. It is doubtful that a significant rearrangement of the dislocations occurs in the time and at the temperature of annealing to which the deformed specimens in these experiments were subjected.[11] Probably some annihilation of the two opposite signs of edge dislocations put in by the bending and straightening of the sample occurs, but the results of the experiments in which the samples were only bent do not differ qualitatively from those for the samples which were straightened as well. Consequently, it is believed that the R *vs* $1/T$ data for the annealed samples shown in Fig. 3 confirm the acceptor action of dislocations in Ge reported by Pearson, Read and Morin.[3,12] Read[6] has advanced the following explanation for the shape of the curves in Fig. 3. As the temperature decreases, the electrostatic repulsion of the charge already trapped by the dislocation line tends to prevent other carriers from accumulating there, and the charge concentration on the line soon becomes insensitive to further decrease in temperature.

The proper identification of the annealable acceptors, on the other hand, is a much more complicated task. There is evidence that lattice vacancies give rise to acceptor action in Ge. Electron bombardment experiments on Ge have been interpreted[13] in terms of the acceptor behavior of vacancies introduced in the crystal lattice by the bombardment.

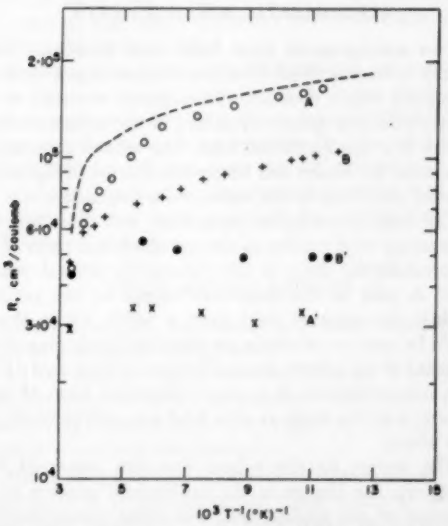


FIG. 3. Hall coefficient *vs* $1/T$ plot for two samples of plastically deformed Sb-doped Ge (A and B) and two control samples (A' and B') which were annealed at temperatures above 600°C for various times (see text). The deformed samples reverted to n-type during annealing. Prior to this they were similar in properties to the samples shown in Fig. 2.

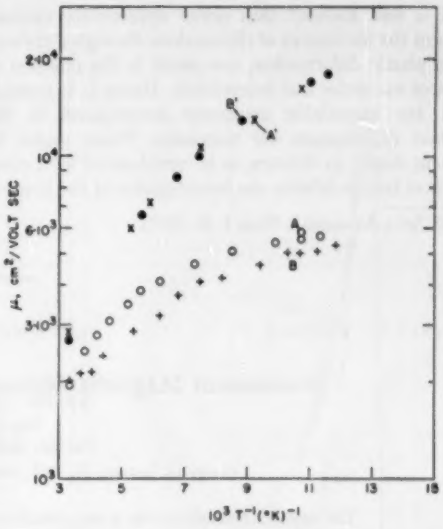


FIG. 4. Hall mobility *vs* $1/T$ for the annealed samples of FIG. 3. Curve labels are as in Fig. 3.

[11] F. L. Vogel, Jr., Acta Metallurgica 3, 95 (1955).

[12] W. T. Read has stated that the temperature at which the samples reported on in reference 3 were deformed was probably somewhat above the 650°C noted there. Considering the relatively light deformation (5-cm radius of curvature) employed by Pearson, Read and Morin, it seems quite likely that their samples were annealed during bending.

[13] Fan, Kaiser, Klontz, Lark-Horovitz, and Pepper, Phys. Rev. 95, 1087 (1954).

It is well known[14] that under appropriate circumstances the movement of dislocations through a crystal, as in plastic deformation, can result in the creation of rows of vacancies and interstitials. Hence it is possible that the annealable acceptors investigated in the present experiments are vacancies. These might be present singly, in clusters, or in combination with other kinds of lattice defects. An investigation of the kinetics of the annealing process would probably help elucidate this important point.

[14] F. Seitz, Advances in Phys. 1, 43 (1952).

ACKNOWLEDGMENTS

The author wishes to thank C. J. Gallagher, with whom some of this work was done, for his continued interest. It is a pleasure to acknowledge the assistance of J. A. Armstrong and R. O. Sisson, who took much of the data.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Permanent Magnetic Moments of a Superconducting Sphere

TOM S. TEASDALE*

The Rice Institute, Houston, Texas

(Received January 31, 1955; revised manuscript received May 2, 1955)

The magnetic flux retained by a polycrystalline sphere of pure tin has been studied by two experimental methods. It is shown that the amount of flux retained by the superconducting sphere exhibits a reproducible dependence on the temperature at which the transition is allowed to occur. The specimen shows a 99.9% Meissner effect for transitions well below the critical temperature, but near to T_c the percent of residual flux increases several hundred-fold. The residual flux is constant as long as the temperature of the specimen is constant, but the flux is not necessarily conserved when the temperature is perturbed.

INTRODUCTION

SIGNIFICANT experiments on superconductivity must now be high-precision measurement on details of the mechanism. The experiments reported here cover several years' work and thousands of data points from two independent experimental methods and are directed toward an investigation of the departures from a complete Meissner effect. In particular, does the residual flux behave as a reproducible function of the path by which the state is reached, or is it subject to random fluctuations?

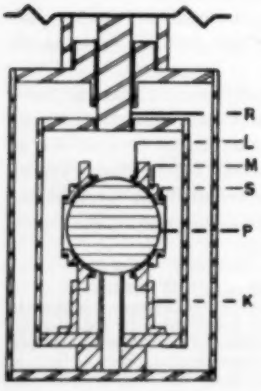


FIG. 1. Drawing of the coil and sphere arrangement. Rotating rod R spins the coil forms K about the sphere P.

The results suggest that the residual flux is a reproducible function in a particular sample, but that it may change when the temperature is changed, even well within the superconducting region.

EXPERIMENTAL ARRANGEMENTS

Two arrangements have been used to obtain the results to be described. The first of these was a torsion pendulum which detected the magnetic moment of a superconducting sphere by means of the torque on the sphere in a small external field. The second apparatus measured the frozen flux by means of a set of flip coils located very close to the superconducting sphere.

The torsion pendulum equipment was described in connection with studies of the gyromagnetic ratio of a superconductor done in this laboratory several years ago.[1] A plot of the reciprocal square of the period against the external field gave a curve whose slope could be used to calculate the effective dipole magnetic moment of the sphere. A small departure from a straight line was attributed to a slight ellipsoidal form of the sphere, but the slope at zero field was independent of this effect.

The torque on the sphere was also measured by observing the change in the equilibrium position as a function of the applied field. For these measurements the sphere was turned so that the axis of the magnetic moment was perpendicular to the external field.

A detail of the flip coils used in the second method is

* Shell Oil Company Fellow in Physics.

[1] Pry, Lathrop, and Houston, Phys. Rev. 86, 905 (1952).

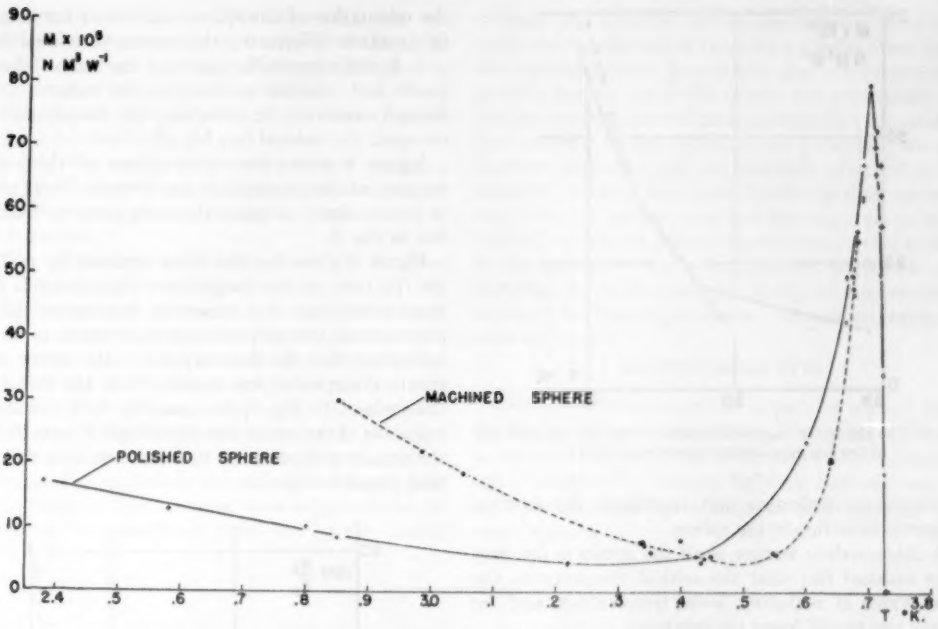


FIG. 2. The magnitude of the magnetic moment of the sphere plotted as a function of the temperature at which the transition occurs. mks units, newton meters² per weber.

shown in Fig. 1. There were two sets of three coils each, and these were rotated together about a central axis through the center of the sphere. The coils marked L, M, and S, had mean diameters of 1.022, 0.820, and 0.590 inches, respectively.

The tin sphere used in this work was cast in a vacuum from spectroscopically pure material. It was machined to have a diameter of 1.0000 plus or minus 0.0004 inches. After a number of measurements on the machined sphere, the amorphous layer on the surface was removed by a method of electropolishing developed by Jacquet.[2] It was later annealed for twelve days at 170°C in an oil bath.

EXPERIMENTAL RESULTS

The data and the conclusions to be drawn from them may be summarized as follows:

1. The residual flux, or residual moment, is a reproducible function of the temperature when the superconducting region is entered by removing the magnetic field at a constant temperature.

Points shown in Fig. 2 give the magnetic moment in mks units as a function of the temperature at which they were produced. They were taken on numerous different occasions over a period of several months. It is clear that the polishing of the surface reduced the residual moment at low temperatures but had little effect near the transition temperature. Figure 3 shows the residual flux measured with the two L coils for the polished sphere and the polished sphere after annealing. The two curves for the annealed spheres represent two orientations of the sphere with reference to the residual field. The quantity B_s of Fig. 3 is computed from the

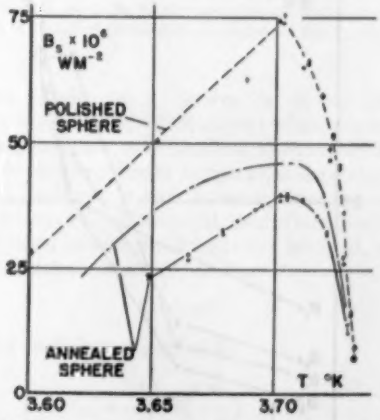


FIG. 3. The average magnetic induction retained by the sphere in webers per square meter. The two curves for the annealed sphere represent two different orientations with respect to field. For clarity, the data points for the middle curve have been omitted.

[2] P. A. Jacquet, International Tin Research and Development Council Bull. 90 (1939).

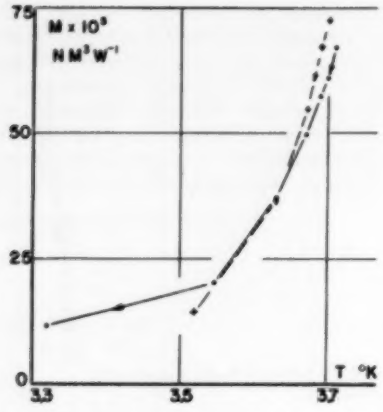


FIG. 4. The change in magnetic moment when the temperature of the superconducting sphere was lowered.

galvanometer deflection and represents the average magnetic induction in the sphere.

A characteristic feature of all the results is the very large residual flux near the critical temperature, the small value at a slightly lower temperature, and the steady rise to still lower temperatures.

It is clear, of course, that the residual moment is associated with imperfections or impurities in the crystal. This work, however, indicates that this association is not a random or accidental one, but is clearly reproducible, that the number and effectiveness of such centers are reduced by annealing, and that after annealing the number of such centers is small enough so that the orientation of the sphere relative to the magnetic field makes a difference in the amount of residual flux.

2. In the superconducting state the residual flux was stable and constant as long as the temperature remained constant. If, however, the temperature was changed, the residual flux fell off.

Figure 4 shows two observations of the residual moment as the temperature was lowered. There seemed to be a tendency to follow the curve given by the stable flux in Fig. 2.

Figure 5 gives the flux measurements by means of the flip coils as the temperature was changed. Again there is evidence of a consistent decrease as the temperature was lowered, although there seems to be some indication that the flux trapped in the center of the sphere disappeared less rapidly than the flux nearer the surface. In Fig. 5, the quantity B_1 is the average induction of the center area of radius 6.22 mm, B_2 is for the annular ring on out to 10.4 mm, and B_3 is the outermost annular ring.

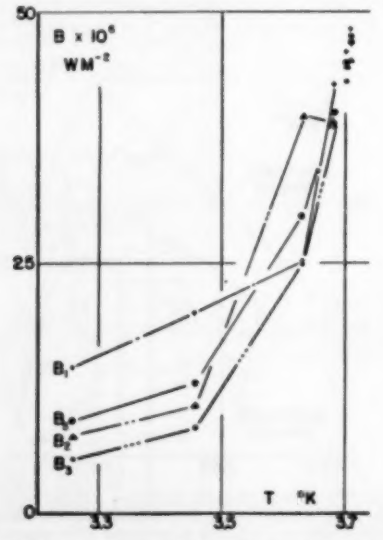


FIG. 5. The change of the average induction when the temperature of the superconducting sphere was lowered.

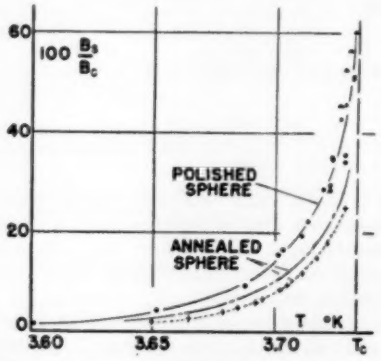


FIG. 6. The percent of flux retained by the sphere between 3.6°K and T_c.

It should be emphasized that the sphere as a whole was well inside the superconducting region at all times when this change was taking place.

3. When the superconducting state was entered near the critical temperature a large residual flux was observed.

Figure 6 indicates the percentage of flux frozen-in under these circumstances; i.e., B_s is the observed magnetic induction remaining in the sphere and B_c is the critical magnetic induction at the particular temperature. Annealing the sphere clearly reduced the proportion of frozen-in flux in accordance with the observations of Pippard,[3] and it is quite possible that continued annealing to more completely remove the physical imperfections might have caused the residual

[3] A. B. Pippard, Third International Conference on Low Temperature Physics and Chemistry (The Rice Institute, Houston, 1953), p. 17. [Also T. E. Faber and A. B. Pippard, *Progress in Low Temperature Physics*, edited by C. J. Gorter (North-Holland Publishing Company, Amsterdam, 1955), Chap. 9.]

flux to disappear entirely. The annealing was not pursued to this extent, since the object of the study was not to produce a perfect Meissner effect but to study the behavior of the residual flux.

4. No effect of the speed of removal of the magnetic field could be observed.

The field was usually removed by simply opening a switch. On other occasions, it was reduced so as to pass through the intermediate state values over a period of several minutes.

5. The residual flux had the direction of the flux from which it was retained within 0.07 radian.

6. No interactions between the residual moment and the over-all surface currents could be observed.

The comparison of the observations with the flip coils and those in which the magnetic moment was measured indicate within the accuracy of these experiments that two superconducting phenomena can take place at the same time. The over-all sphere was superconducting and rotated in the external magnetic field without damping. Hence, there were supercurrents on the surface, and inside of these the sphere moved without any apparent interaction between these currents and the material of the sphere. On the other hand, the residual moment is evidently due to supercurrents, possibly located inside the sphere, but presumably extending to the surface at those points where the residual flux emerges. These currents are persistent, do not decrease with time, and are definitely attached to the material lattice of the sphere. Evidently these currents turn when the sphere turns, and there appears to be no interaction with the general supercurrents on the surface of the sphere. It was never possible to observe any damping due to the presence of the residual moment, although in these experiments the residual moments were very small.

ACKNOWLEDGMENTS

This project was supported in part by a grant from the National Science Foundation. The author is pleased to express his appreciation to Professor W. V. Houston and Professor C. F. Squire for their guidance and to his colleagues who gave time in assisting with the measurements.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Ferromagnetic Hall Coefficient of Nickel Alloys

ALBERT I. SCHINDLER AND EDWARD I. SALKOVITZ
United States Naval Research Laboratory, Washington, D. C.
(Received April 21, 1955)

Published Hall coefficient data for binary Ni alloys have been examined in light of the Karplus-Luttinger theory of the ferromagnetic Hall effect. This theory indicates that the extraordinary Hall coefficient is related to the square of the electrical resistivity, $R_1=A\rho^2$. Neither the exponent nor the coefficient of ρ varies as additions of up to 30 atomic percent Cu are made to pure Ni, and as the temperature is varied from 20°K to 290°K. On the other hand, small additions of either Al, Si, or Sn to Ni cause a significant variation in A, but appear to leave the exponent of ρ unaltered. A tentative explanation is offered for the difference in behavior.

IN ferromagnetic materials, the Hall electric field may be written in the following manner:

$$e_H=R_0H+R_1M=R_0(H+4\pi\alpha M);\quad R_1=4\pi\alpha R_0. \quad (1)$$

Recently, Karplus and Luttinger[1] separated R_1 into two parts, one, $(4\pi R_0)$, arising from an internal field due to M, the other, (R_1'), arising as a direct consequence of the spin-orbit interaction. R_1 was then written as:

$$R_1=R_1'+4\pi R_0. \quad (2)$$

At all temperatures except the very lowest, $4\pi R_0 \ll R_1'$: therefore R_1 and R_1' are essentially equal and are considered so in this discussion. Karplus and Luttinger further showed that the large magnitude of R_1' is a consequence of the spin-orbit interaction of the $3d$ electrons. Their theory derives as an equilibrium property of the metal, a Hall current which arises from a net magnetization, a longitudinal electric field, and a spin-orbit coupling. Strong temperature dependence of the Hall coefficient is then introduced by converting this Hall current to a Hall electric field of standard form. It is the latter process which leads to a factor ρ^2, where ρ is the electrical resistivity,[2] so that

$$R_1'=A\rho^2, \quad (3)$$

where the coefficient A is given by

$$A=\frac{2e^2H_{so}}{m\Delta^2}\delta\left\langle\frac{m}{m^*}\right\rangle_{d\omega}.$$

Here e and m are the electronic charge and mass, H_{so} is

[1] R. Karplus and J. M. Luttinger, Phys. Rev. **95**, 1154 (1954).

[2] See Eqs. (2.20), (2.21), and (2.22) in reference 1.

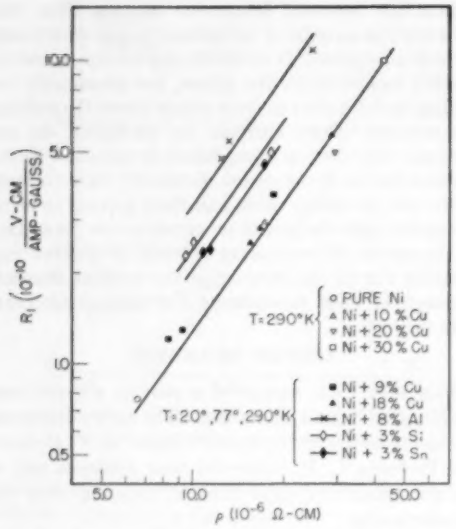


FIG. 1. Extraordinary Hall coefficient as a function of resistivity for Ni alloys.

the magnetic field arising from the spin-orbit interaction, δ is the number of unfilled $3d$ sub-bands, Δ^2 is the mean square energy separation between different $3d$ sub-bands for the same k at the top of their Fermi surfaces and $\langle m/m^* \rangle_{3d}$ is the ratio of the electronic mass to the effective mass of the $3d$ electrons averaged over the $3d$ sub-bands. Measurements[2] of R_1 and ρ at various temperatures for Fe and Ni confirm a power law resistivity dependence, although the exponent of ρ was found to be 1.94 for Fe and 1.42 for 99.5% pure Ni.

Kooi[4] showed that alloying Fe with up to 5% Si produced no change in either A or the exponent of ρ. There are also data available on the Hall coefficient of binary alloys of Ni containing Cu, Al, Si, or Sn.[5,6] Schindler and Pugh gave room temperature data for R_0 and α for Ni—Cu alloys, from which R_1 can be calculated. For these alloys, the variation in ρ due to alloying was obtained from the work of Svensson.[7]

The data have been plotted in Fig. 1 on a log-log plot. The solid line passing through the Ni—Cu data is a reproduction of the curve for the temperature variation of R_1 and ρ for 99.5% pure Ni.[3] The close fit of the Ni—Cu data to the 99.5% pure Ni line indicate that neither A nor the exponent of ρ change in the Ni—Cu system, up to 30 atomic percent copper, and over a temperature range from 20°K to 290°K. The three other Ni alloys in Fig. 1 show approximately the same exponent of ρ, but different values of A. Table I tabulates A, assuming $R_1 = A\rho^{1.42}$. Lavine[8] finds that the exponent of ρ approaches a value of 2 for high purity nickel. However, the data of Smit and Volger[6] for carbonyl nickel at 77°K and 290°K indicate a slope of less than 1.4.

TABLE I. Values of the parameter A for Ni alloys.

Composition	A	References
Ni (99.5% pure)	1.61×10^{-3}	3
Ni—Cu (to 30%)	1.61×10^{-3}	5, 6
Ni—8% Al	4.21×10^{-3}	6
Ni—3% Si	2.99×10^{-3}	6
Ni—3% Sn	2.57×10^{-3}	6

The coefficient A contains band-dependent parameters δ, Δ^2, and $\langle m/m^* \rangle_{3d}$, plus a spin-orbit parameter H_{so}. It is quite probable that δ and Δ^2 are not greatly affected by alloying.[9] However as the $3d$ band fills $\langle m/m^* \rangle_{3d}$ (where the average is over all the filled states), might be expected to change with electron concentration, but not very much since the $3d$ band is almost full in pure Ni; in the case of dilute alloys the changes would be particularly small. But since A does change in these alloys, the burden of the variation in A must lie in H_{so}. On the other hand, additions up to 30 percent Cu do not sensibly alter A. This suggests that no significant changes occur in H_{so} in these Ni—Cu alloys. Since Cu and Ni are adjacent members of the periodic table, this is not surprising, for H_{so} is closely related to the effective atomic number.

It should also be pointed out that Smit and Volger give data for the temperature dependence of the R_1 and ρ for alloys of Ni with various transition metals. In these cases, further complications arise since both the coefficient A and the exponent of ρ change as a function of alloying addition.

We would like to acknowledge helpful discussions with Dr. P. Marcus, and would like to thank Dr. J. Smit for showing us some of his unpublished data.

[3] J. P. Jan, Helv. Phys. Acta **25**, 677 (1952); J. P. Jan and J. M. Gijsman, Physica **18**, 339 (1952); E. H. Butler and E. M. Pugh, Phys. Rev. **57**, 916 (1940).
[4] C. Kooi, Phys. Rev. **95**, 843 (1954).
[5] A. I. Schindler and E. M. Pugh, Phys. Rev. **89**, 295 (1953).
[6] J. Smit and J. Volger, Phys. Rev. **92**, 1576 (1953).
[7] R. Svensson, Ann. Physik 5, **25**, 263 (1936).
[8] J. Lavine, Harvard University (private communication).
[9] G. C. Fletcher, Proc. Phys. Soc. **A65**, 192 (1952).

Hyperfine Structure in the Inversion Spectrum of $N^{14}H_3$ by a New High-Resolution Microwave Spectrometer*

J. P. Gordon†
Columbia University, New York, New York
(Received May 4, 1955)

The hyperfine structure of the inversion spectrum of $N^{14}H_3$ has been reexamined with an ultra-high resolution spectrometer. Lines whose total width at half-maximum is seven kc-sec have been obtained in this spectrometer. Such narrow lines allow resolution of magnetic hyperfine structure caused by reorientation of the spins of the hydrogen nuclei. The new structure has been satisfactorily explained to within the experimental error of about 1 kc/sec by considering the various interactions of the magnetic moments of the hydrogen nuclei with the molecular fields. The interaction energy of the nitrogen nucleus has been remeasured with higher resolution than was possible previously, and indications were found that $|eqQ|$ for the lower inversion state is larger by about 0.01% than that for the upper state. The sign of eqQ was directly determined to be negative. A theoretical treatment of the magnetic interactions has been made which is slightly different from that which has previously been used. The method allows some simplification of the form of the magnetic interactions.

INTRODUCTION

THE inversion spectrum of the symmetric top molecule $N^{14}H_3$ has been the subject of a great deal of study ever since it was first observed by Cleeton and Williams in 1934.[1,2] Hyperfine structure in this spectrum due to quadrupole interaction of the nitrogen nucleus was first noticed in 1946 by Good.[3] Quite recently, a thorough theoretical and experimental study of the hyperfine structure of this spectrum has been carried out by Gunther-Mohr *et al.*[4,5] which explained a doubling of the $K=1$ lines in terms of the magnetic interactions of the hydrogen nuclei with the molecular magnetic fields.

A significant increase in experimental resolution afforded by the molecular beam spectrometer described in this paper has made it desirable to remeasure once again in greater detail the hyperfine structure of this spectrum. Previously undetected structure caused by the reorientation of the spins of the hydrogen nuclei has been observed and theoretically explained to within experimental error. The theoretical analysis of Gunther-Mohr *et al.* has been extended to include evaluation of the mutual spin-spin interaction of the three hydrogen nuclei, and simplification of some of the terms of the Hamiltonian has been effected through the use of the symmetry properties of the molecule.

* Work supported jointly by the Signal Corps, the Office of Naval Research, and the Air Research and Development Command.

† Now at the Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

[1] C. E. Cleeton and N. H. Williams, Phys. Rev. **46**, 235 (1934).

[2] The reader is referred to C. H. Townes and P. Kisliuk, National Bureau of Standards Circular, 518, June 23, 1952 (unpublished), in which will be found a summary of the experimental and theoretical work on this molecule together with a survey of the literature and a tabulation of microwave transition frequencies.

[3] W. E. Good, Phys. Rev. **70**, 213 (1946).

[4] Gunther-Mohr, White, Schawlow, Good, and Coles, Phys. Rev. **94**, 1184 (1954).

[5] Gunther-Mohr, Townes, and Van Vleck, Phys. Rev. **94**, 1191 (1954).

THEORY

The experimental[4] and theoretical[5] papers by Gunther-Mohr *et al.* form the background for the following work. For convenience, they will be referred to as GM I and GM II respectively.

The rotational state of the molecule in zero field will be specified by the following quantum numbers:

- $\mathbf{J}$ total molecular angular momentum excluding nuclear spins.
- K projection of $\mathbf{J}$ on the molecular axis of symmetry.
- $\mathbf{I}_N$ spin of the nitrogen nucleus.
- $\mathbf{I}_i$ ($i=1, 2, 3$) spin of the ith hydrogen nucleus.
- $\mathbf{I}$ sum of the spins of the three hydrogen nuclei.
- $\mathbf{F}_1$ sum of $\mathbf{J}$ and $\mathbf{I}_N$ (this is a good quantum number because the quadrupole interaction energy of the nitrogen is considerably larger than the magnetic interaction energy of the hydrogens).
- $\mathbf{F}$ sum of F_1 and $\mathbf{I}$.
- M_F projection of F on some axis fixed in space.

The calculation of the various matrix elements will be made in the molecular frame coupling scheme, as described by Van Vleck[6] and GM II. The molecular frame scheme is based on the fact that if one refers all angular momenta to set of axes fixed in the molecule, then $\mathbf{F}$, $\mathbf{I}_N$, and $\mathbf{I}$ form a set of commuting angular momenta, just as in the laboratory frame, $\mathbf{J}$, $\mathbf{I}_N$, and $\mathbf{I}$ form a commuting set. In the molecular frame, however, the commutation relations of $\mathbf{F}$ and $\mathbf{F}_1$ change sign: i.e., $[F_x,F_y]=-iF_z$; where x, y, and z now refer to axes fixed in the molecule. The "internal" angular momenta $\mathbf{I}$ and $\mathbf{I}_N$, however, retain the old commutation relations. If one reverses the signs of these internal angular momenta, setting $\mathbf{I}_N'=-\mathbf{I}_N$, $\mathbf{I}'=-\mathbf{I}$, then all of the angular momenta commute with the anomalous sign, and one has the molecular frame coupling scheme

$$\mathbf{F}_1=\mathbf{F}+\mathbf{I}', \quad \mathbf{J}=\mathbf{F}_1+\mathbf{I}_N',$$

[6] J. H. Van Vleck, Revs. Modern Phys. **23**, 213 (1953).

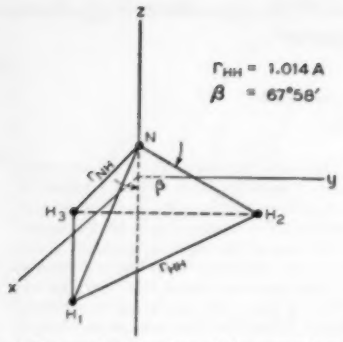


FIG. 1. Geometrical structure of the ammonia molecule.

and such relations as

$$\mathbf{I}'^2=\mathbf{I}^2=I(I+1), \quad \mathbf{F}\cdot\mathbf{I}'=-\mathbf{F}\cdot\mathbf{I}.$$

The matrix elements of Condon and Shortley[7] may be used for this coupling scheme, provided one reverses the signs of all the imaginary ones. The molecular frame has a distinct advantage, for calculations, in that the positions of the nuclei are fixed in the coordinate system (except for vibrations) and it becomes unnecessary to know the matrix elements of the direction cosines which determine the angular position of the molecule in space.

The interactions which are important in the hyperfine structure of the inversion spectrum of ammonia are listed in Table I.[8] They include the effects previously evaluated in GM II, with the addition of the mutual spin-spin interaction of the magnetic moments of the hydrogen nuclei. The reader is referred to GM I and GM II for the evaluation of any constants not explicitly defined here. The combined effect of these interactions satisfactorily fits almost all the experimental data to within the accuracy of measurement. The few apparent discrepancies which still remain will be discussed later.

We shall proceed to a calculation of the hydrogen spin-spin interaction. Reference should be made to Fig. 1 which shows the structure of the molecule and the positions of the nuclei with respect to the molecular axes. The z axis is the axis of symmetry of the molecule, the xz plane contains hydrogen No. 1, and the projection of the position of hydrogen No. 2 on the xy plane lies in its second quadrant. The origin of the coordinate system is taken at the center of gravity of the molecule. The x, y, and z axes form a right-handed set of coordinates.

[7] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge University Press, London, 1951).

[8] Some errors noticed in GM I and GM II are corrected here and in Eq. (5). These errors include: (*a*) The function f of Eq. (5) is written incorrectly in GM II. (*b*) in GM I and II, the magnetic constants α and ρ should be replaced in all equations by $\alpha/2$ and $\rho/2$. These constants, of course, keep the same definitions as before. (*c*) in GM II, Appendix III, Eq. (A23), the proportionality constant $\mathfrak{D}_1(J,I_N)/[J(J+1)]$ should be replaced by $\varphi(J,I_N')/[J(J+1)]$ where $\varphi(J,I_N')=2\mathfrak{D}_2(J,I_N)-\mathbf{I}_N\cdot\mathbf{J}$.

The energy of the spin-spin interactions among the three hydrogen nuclei is given by[9]

$$G_1^{HH}=(g_H\mu_0)^2 r_{HH}^{-3} \times\sum_{i<j}[\mathbf{I}_i\cdot\mathbf{I}_j-3r_{HH}^{-2}(\mathbf{I}_i\cdot\mathbf{r}_{ij})(\mathbf{I}_j\cdot\mathbf{r}_{ij})], \quad (1)$$

where r_{HH} is the distance between hydrogen nuclei, and $\mathbf{r}_{ij}$ is the vector distance between hydrogen nuclei i and j. As shown in the Appendix, the only matrix elements of this interaction which affect the spectrum in an observable way are those which are diagonal in K, and if only these elements are considered, it is shown there that the interaction can be written in the equivalent form:

$$G_1^{HH}=-\tfrac{1}{4}(g_H\mu_0)^2 r_{HH}^{-3}(\mathbf{I}^2-3\mathbf{I}_z^2), \quad (2)$$

where I_z is the projection of I on the molecular symmetry axis. This term may then be evaluated in a similar way to that by which the term $(\mathbf{I}_{Nz}^2-\mathbf{I}_N^2/3)$ is evaluated in GM II, giving the result:

$$G_1^{HH}=-\tfrac{1}{4}(g_H\mu_0)^2\langle v|r_{HH}^{-3}|v\rangle \times\frac{\varphi(F_1,I')\varphi(J,F_1)}{F_1(F_1+1)}\left(1-\frac{3K^2}{J(J+1)}\right), \quad (3)$$

TABLE I. Interactions important in the hyperfine structure of the inversion spectrum of $N^{14}H_3$.

Energy term	Physical interpretation
$-eqQ\left(1-\frac{3K^2}{J(J-H)}\right)\mathfrak{D}_1(J,I_N)$	Nitrogen quadrupole interaction
$\left[a+\frac{(b-a)K^2}{J(J+1)}\right](\mathbf{I}_N\cdot\mathbf{J})$	Nitrogen magnetic interaction with molecular rotation
$\left[A+\frac{cK^2}{J(J+1)}-\delta_{K1}(-1)^{J+v}B\right](\mathbf{I}\cdot\mathbf{J})$	Hydrogen magnetic interaction with molecular rotation
$g_H g_N\mu_0^2 r_{NH}^{-3}\sum_{i=1,2,3}[(\mathbf{I}_i\cdot\mathbf{I}_N) -3(\mathbf{I}_i\cdot\mathbf{r}_{iN})(\mathbf{I}_N\cdot\mathbf{r}_{iN})]$ or $\mathfrak{D}_1(\mathbf{I}\cdot\mathbf{I}_N-3I_zI_{Nz})+3\mathfrak{D}_2(I_xI_{Nx}+I_yI_{Ny}-I_{Iz}I_{Nz})$	Hydrogen-nitrogen spin-spin interaction
$(g_H\mu_0)^2 r_{HH}^{-3}\sum_{i<j}[\mathbf{I}_i\cdot\mathbf{I}_j-3(\mathbf{I}_i\cdot\mathbf{r}_{ij})(\mathbf{I}_j\cdot\mathbf{r}_{ij})]$ or $-\tfrac{1}{4}\mathfrak{D}_1(I^2-3I_z^2)$,	Hydrogen-hydrogen spin-spin interaction

where

$$\mathfrak{D}_1=g_H g_N\mu_0^2\langle v|r_{NH}^{-3}(1-\tfrac{3}{4}\sin^2\beta)|v\rangle,$$
$$\mathfrak{D}_2=g_H g_N\mu_0^2\langle v|r_{NH}^{-3}\tfrac{3}{4}\sin^2\beta|v\rangle,$$
$$\mathfrak{D}_3=(g_H\mu_0)^2\langle v|r_{HH}^{-3}|v\rangle,$$
$$A=\tfrac{1}{2}(\alpha+\rho), \quad B=\tfrac{1}{2}(\rho-\alpha), \quad C=\gamma-\tfrac{1}{2}(\alpha+\rho);$$

α, ρ, and γ are proportional to the magnetic fields at the position of hydrogen nucleus 1 (see Fig. 1) caused by rotations of the molecule about the x, y, and z axes, respectively (see GM II).

[9] The notation here for the Hamiltonian is patterned after GM II.

where $\langle v|r_{HH}^{-3}|v\rangle$ indicates that r_{HH}^{-3} is to be averaged over the inversion, and

$$\varphi(A,B)=\frac{6(\mathbf{A}\cdot\mathbf{B})^2-3(\mathbf{A}\cdot\mathbf{B})-2A(A+1)B(B+1)}{(2A+3)(2A-1)}. \quad (4)$$

It can be shown from symmetry arguments that the total hydrogen spin depends on K, so that when K is a multiple of 3, I is $\frac{3}{2}$, while for all other values of K, $I=\frac{1}{2}$. Since states with $I=\frac{1}{2}$ must have $I_z^2=\frac{1}{4}$, it can be seen from Eq. (2) that G_1^{HH} is zero in this case. Thus this interaction does not affect the hyperfine structure of the energy levels unless K is a multiple of three. When K is a multiple of three, however, it forms an important component of the magnetic structure. In computing the magnitude of the coupling constant $(g_H\mu_0)^2\langle v|r_{HH}^{-3}|v\rangle$ we have used for $\langle v|r_{HH}^{-3}|v\rangle$ merely the inverse cube of the equilibrium value of r_{HH} which was obtained using the values given by Herzberg[10] for the hydrogen nitrogen distance r_{HH} and the apex angle β which are illustrated in Fig. 1. These values are $r_{HH}=1.014\times10^{-8}$ cm, $\beta=67°58'$. The coupling constant was thus calculated to be 27.7 kc/sec, and this value provides a good fit to the experimental spectrum.

Combining this result with those of GM II, we arrive at the final form of the complete hyperfine energy. This equation is valid, as explained below, so long as the quadrupole energy term is large compared to the magnetic terms.

$$\begin{aligned}W_{JJKF_1F}=&-\langle v|eQq|v\rangle\left(1-\frac{3K^2}{J(J+1)}\right)\Omega_1(J,I_N)\\&+\left[a+\frac{(b-a)K^2}{J(J+1)}\right](\mathbf{I}_N\cdot\mathbf{J})\\&+\left[A+\frac{CK^2}{J(J+1)}\right]\frac{(\mathbf{I}\cdot\mathbf{F}_1)(\mathbf{F}_1\cdot\mathbf{J})}{F_1(F_1+1)}\\&+2g_Hg_N\mu_0^2\langle v|r_{HH}^{-3}(1-\tfrac{3}{2}\sin^2\beta|v\rangle\\&\times\frac{(\mathbf{F}_1\cdot\mathbf{I})\Omega_2(J,I_N)}{F_1(F_1+1)}\left(1-\frac{3K^2}{J(J+1)}\right)\\&-\tfrac{1}{4}(g_H\mu_0)^2\langle v|r_{HH}^{-3}|v\rangle\frac{\varphi(F_1,I')\varphi(J,F_1)}{F_1(F_1+1)}\\&\times\left(1-\frac{3K^2}{J(J+1)}\right)-\delta_{K1}(-1)^{J+\frac{1}{2}}2\mathbf{F}, \quad (5)\end{aligned}$$

where

$$\begin{aligned}2f=&\langle v|r_{HH}^{-3}\tfrac{3}{2}\sin^2\beta|v\rangle 2g_Hg_N\mu_0^2\\&\times\frac{(\mathbf{F}_1\cdot\mathbf{I})\Omega_2(J,I_N)}{F_1(F_1+1)}+B\frac{(\mathbf{I}\cdot\mathbf{F}_1)(\mathbf{F}_1\cdot\mathbf{J})}{F_1(F_1+1)},\end{aligned}$$

[10] G. Herzberg, *Infra-Red and Raman Spectra* (D. Van Nostrand and Company, Inc., New York, 1945).

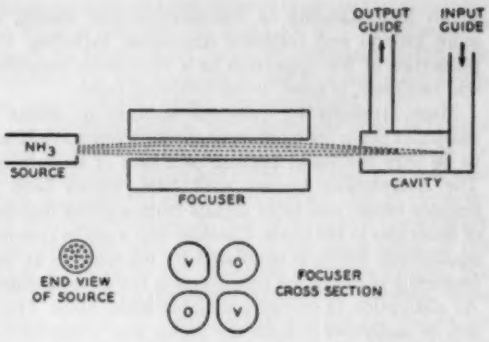


FIG. 2. Block diagram of the molecular beam oscillator.

$$\Omega_1(J,I_N)=\frac{3(\mathbf{I}_N\cdot\mathbf{J})^2+\frac{3}{2}(\mathbf{I}_N\cdot\mathbf{J})-I_N(I_N+1)J(J+1)}{(2J-1)(2J+3)2I_N(2I_N-1)},$$

$$\Omega_2(J,I_N)=\frac{3(\mathbf{I}_N\cdot\mathbf{J})^2+2J(J+1)(\mathbf{I}_N\cdot\mathbf{J})-I_N(I_N+1)J(J+1)}{(2J-1)(2J+3)}.$$

In addition to this expression, the magnetic terms of Table I have matrix elements off diagonal in F_1 which might be expected to contribute measurable shifts in some of the lines where the quadrupole splittings are small. For the 3–2 line, in particular, the factor $[1-3K^2/J(J+1)]$ is equal to zero, so that the quadrupole coupling energy disappears, F_1 is no longer a good quantum number, and expression 5 is no longer valid. Of the lines we measured, the largest effect of terms off-diagonal in F_1 appears in the 4–3 line, since for this line the quadrupole splittings were small and the magnetic energies large. In this case some of the individual hyperfine structure components are shifted by as much as two kc/sec, but this did not produce a detectable change in the envelope of all components of the lines.

EXPERIMENTAL CONSIDERATIONS

The experimental results were obtained by use of an ultra-high resolution spectrometer.[11] A block diagram of this spectrometer is shown in Fig. 2. Ammonia gas at room temperature is maintained in the source at 5 to 20 mm pressure. Molecules emerge from the source through a group of fine tubes and travel through an electrostatic focuser in which a highly nonlinear electric field focuses the molecules which are in upper inversion states. Molecules in lower inversion states are defocused and so removed from the beam. The beam then traverses a high-Q resonant cavity tuned to an inversion frequency, and there transitions are induced by microwave radiation from a 2K50 klystron. We shall discuss here only those properties of the apparatus which are necessary

[11] Gordon, Zeiger, and Townes, Phys. Rev. 95, 282 (1954).

to an understanding of the spectroscopic results. A more general and complete discussion, including the properties of the apparatus as a microwave amplifier and oscillator, is given in the following paper.

Most attempts by previous workers to utilize a molecular beam for microwave spectroscopy have failed to be very successful because of a lack of sensitivity. The electrostatic focuser used here, allows both a sizeable beam, and large signals from a given number of molecules in the beam. Consider first a gas in thermal equilibrium which is irradiated by microwaves at the frequency of one of the transitions of the gas molecules. An absorption of energy invariably takes place. There will be molecules in both the higher and lower energy states of the transition, and the microwaves will excite transitions both up and down with equal probability. The total rate of upward transitions will, however, be slightly larger than the rate of downward transitions, since the Boltzmann distribution gives the lower state a population larger than the upper by approximately $N(h\nu/kT)$, where N is the total number of molecules in the lower state. For $T=300°$K, $kT/10^{8}$h equals 6.63×10^{6} Mc/sec, and so $h\nu/kT$ is approximately 1/250 at the frequency of 1.25-cm microwaves. At this frequency, then, the net effect is produced by only 1/250 of the molecules which are capable of absorbing energy. In the present apparatus, only molecules in upper inversion states reach the cavity. Since they can all contribute energy to the exciting microwave field, there results a substantial improvement in the sensitivity of the apparatus. The fact that the upper inversion states are focused means that the molecular response to the exciting radiation is an emission of energy, so that emission lines are observed rather than absorption lines.

Cylindrical cavities were used, which resonated in the TE_{011} mode at the frequency of interest. In order to provide a long interaction region and so to cut down the line width (this is discussed in detail in the following paper), the cavities were made approximately four inches in length. They were tuned by means of short sections of enlarged diameter and variable length at one end. The beam entered through a circular hole in the other end. In order to make cavities of this length for the TE_{011} mode, their diameters had to be very close to the cut-off values. This restricted the tuning range of each cavity to about 50 Mc/sec, and so a new cavity usually had to be constructed for each line which was examined. The cavities were made of copper and had values of Q near 12 000. Suprisingly little radiation from the beam entrance hole was encountered, even though the hole diameter was two thirds of the diameter of the cavity.

Emission lines whose total width at half-maximum is seven kc/sec have been obtained, and it is estimated that from two to four kc/sec of this width is due to spurious frequency modulation of the microwave signal from the 2K50 klystron which was used to induce the transitions, rather than to the actual width of the spectral lines.

The field of the focuser exerts forces on the molecules due to their induced dipole moments. In zero field the ammonia molecule has no average dipole moment because of the inversion, while with increasing field, the inversion is slowly quenched and an average dipole moment appears. If hyperfine effects are neglected the energies of the inversion states may be written for all field strengths as

$$W=W_0\pm\left[\left(\frac{h\nu_0}{2}\right)^2+\left(\mu\mathcal{E}\frac{MK}{J(J+1)}\right)^2\right]^{\frac{1}{2}}, \qquad (6)$$

where W_0 is the average energy of the upper and lower inversion levels, ν_0 is the inversion frequency in zero electric field, μ is the permanent dipole moment the molecule would have if the inversion did not occur, $\mathcal{E}$ is the magnitude of the electric field, and M is the projection of J on the direction of the field. Since the changes in magnitude and direction of the field the molecule sees as it travels through the focuser occur very slowly compared to the rate of precession of J around the field, J maintains its orientation with respect to the field direction during the whole flight of of the molecule. Therefore M is a constant of the motion.

The focuser, as shown in Fig. 2, is composed of four cylindrical electrodes whose inner faces are shaped to form hyperbolae. Two opposing electrodes are maintained at a high voltage V, while the other two are kept at ground. This arrangement produces an electric potential in the gap between the electrodes of the form $\varphi=\varphi_0 r^2\cos2\theta$; where r and θ are cylindrical coordinates of a system whose axis is the axis of the focuser. The magnitude of the electric field produced by this potential is proportional to r. The force exerted on the molecules by this field is, from (6),

$$f_r=-\frac{\partial W}{\partial r}=\mp\frac{[\mu MK/J(J+1)]^2\mathcal{E}d\mathcal{E}/dr}{\{(h\nu_0/2)^2+[\mu\mathcal{E}MK/J(J+1)]^2\}^{\frac{1}{2}}}. \qquad (7)$$

It is radial and proportional to r for small field strengths. The force is directed inward for molecules in upper inversion states (negative sign) and outward for molecules in lower inversion states (positive sign).

Under the assumption of small electric fields, the number of molecules in the upper state and with a given value of M which are held in the focuser by the restoring force is approximately proportional to M^2. This can be seen as follows. Assume that the source consists of a small hole located on the axis of the focuser from which molecules effuse in all directions. Assume further that all molecules which can travel farther away from the axis of the focuser than a distance of r_0 collide with the focuser electrodes and are lost from the beam. It can easily be shown that all molecules whose initial velocity perpendicular to the axis of the focuser is less than a

given v_{m0} are kept within the focuser, where v_{m0} is proportional to M. For any axial velocity v_z, all molecules in a given M state which emerge from the source within a solid angle $\Omega_M=\pi(v_{m0}/v_z)^2$ are thus kept within the focuser, and since v_{m0} is proportional to M, Ω_M is proportional to M^2. Since Ω_M is small, the number of molecules in solid angle Ω_M is proportional to Ω_M, and thus to M^2. Actually none of the assumptions made here is exactly correct, the principal deviation being due to the decrease in polarizability of the molecule at strong fields. However, calculations show that the results are not changed by this effect in a very important way. Hence we shall assume for simplicity, that the field is relatively weak ($\mu\mathcal{E}\ll h\nu_0$), so that the number of upper state molecules trapped by the potential minimum of the focuser, and thus delivered to the cavity, is proportional to M^2.

We shall see that as the molecules enter, traverse, and leave the focuser, very few transitions are made between the various quadrupole levels of a given inversion state. This fact plus the dependence of the focusing strength on M will be shown to have the effect of substantially increasing the relative populations of the higher energy quadrupole levels inside the cavity. The quadrupole satellites on the high-frequency side of the main line are made relatively strong by this effect.

Neglecting for a moment the effects of the hydrogen spins, we see that in small electric fields, J and the nitrogen spin I_N are coupled together by the quadrupole interaction and the stationary states of the molecule may be represented by the quantum numbers J, I_N, F_1, and M_F, where $\mathbf{F}_1=\mathbf{J}+\mathbf{I}_N$ and M_F is the projection of $\mathbf{F}_1$ along the direction of the field. At high fields, $\mathbf{J}$ and $\mathbf{I}_N$ are decoupled and separately quantized along the field direction. At all intermediate fields, the stationary states have a one to one correspondence with the low- or high-field states. If a molecule exists at low field in a particular quantum state and then the field is increased infinitely slowly (adiabatic approximation) the state of the molecule will progress through the intermediate configurations and at high field will be found in a particular high-field state. If the field is turned on within a finite time, then transitions will be possible, and the molecule may then be found in any of several high-field states.

The criterion for whether such transitions will take place is discussed in most textbooks on quantum mechanics. Let $\hbar\omega_{km}$ be the energy difference between states k and m. Then if there is only a small fractional change in $\hbar\omega_{km}$ in a time given by $1/\omega_{km}$, few transitions between states k and m are to be expected. The most probable velocity of the ammonia molecules in the beam is about 4×10^4 cm/sec and the distance they travel in entering or leaving the focuser is about one centimeter, so that they take about 2.5×10^{-5} second to enter the field. The quadrupole level splittings are of the order of 1.5 Mc/sec, so that for transitions to occur between these levels, the energy differences between quadrupole states must undergo large fractional changes in times of the order of $[2\pi\times1.5\times10^6]^{-1}$ or about 10^{-7} sec. This time is short, however, compared to the time required for the molecule to enter or leave the fields, so that we should expect few transitions.

In addition, each of the quadrupole levels is further split in zero field by the magnetic interactions of the hydrogen spin, which are listed in Table I. These zero field magnetic hyperfine splittings are only of the order of 50 kc/sec, however, so that in the case of these energy splittings the time $1/\omega_{km}$ is of the same order as the time of entry of the molecules into the focuser. We should thus expect many transitions between these levels, and in our computing theoretical line shapes, we have assumed that in the cavity all the magnetic hyperfine sublevels of a given quadrupole level are equally populated.

As an example of the intensity shifts produced by adiabatic focusing we shall take the case of the 3–3 line, which is one of those experimentally measured. Since eqQ is negative, Eq. (5) shows that the $F_1=3$ state has the highest energy, followed by the $F_1=4$ and 2 states respectively. Of the seven low-field M_F states with $F_1=3$, four go at high fields to $|M|=3$ states, two to $|M|=2$, and one to $|M|=1$. The other low-field states are somewhat similarly distributed among the high-field states. However, there is a significant trend for low-energy low-field states to go predominantly into the low-energy high-field states. Following our assumption that the number of molecules in a given M state which reach the cavity is proportional to M^2, the population of each F_1 level in the cavity must be weighted according to the factor $(2F_1+1)^{-1}\sum_M N(M)M^2$, where $N(M)$ is the number of low-field states which go into the particular high-field state with quantum number M. For the 3–3 line, these weighting factors are 45/7, 37/9, and 2/5 respectively for the states $F_1=3$, 4, and 2. The high-energy quadrupole levels thus have substantially greater populations in the cavity. The relative intensities of the transitions originating from the various quadrupole levels must be weighted by these factors.

The effect of this varying of the weight-factor was experimentally observed as a variation in intensity of the transitions. Under conditions of thermal equilibrium, the hyperfine pattern of the 3–3 line should be almost exactly symmetrical intensitywise, and in particular the two transitions $F_1=3\rightarrow2$ and $F_1=2\rightarrow3$ equal in intensity. Actually, while the $F_1=3$ to 2 quadrupole satellite was quite strong (see Fig. 4), the $F_1=2$ to 3 quadrupole satellite on the low frequency side of the main line was just barely observable.

The relative intensities of the quadrupole satellites were used to provide a direct check on the sign of eqQ. Under the correct assignment of a negative eqQ, both quadrupole satellites on the high frequency side of the main line originate from the same quadrupole level and thus their relative intensities are unchanged by the

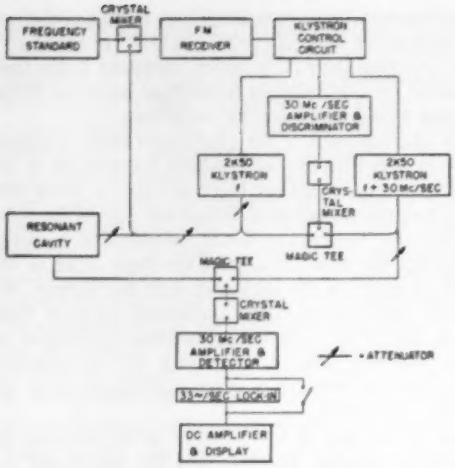


FIG. 3. Block diagram of the heterodyne detection system used in the experimental work.

adiabatic focusing, while the quadrupole satellites on the low frequency side originate from different quadrupole levels, and so have different intensities. The $K=1$ line also checked the sign of eqQ, since the details of its asymmetric hyperfine structure depend on the relative sign of eqQ and the nuclear spin-spin interaction. The sign of the latter interaction is of course known from the nuclear magnetic moments. This determination also shows that eqQ is negative.

The large changes in the relative populations of the quadrupole levels due to focusing also allow the possibility of producing direct quadrupole transitions in the beam in the region between the focuser and the cavity. The cavity would then be used as a detector, the quadrupole resonances being observed by the changes in intensity of the quadrupole satellites of the microwave transitions.

ELECTRONICS AND MEASUREMENT TECHNIQUE

A block diagram of the heterodyne detection system used in this work appears in Fig. 3. The carrier power from the signal klystron $K1$ plus power emitted from the beam is mixed in a 1N26 crystal with power from a second klystron $K2$ which is maintained by an automatic frequency control AFC circuit at a frequency 30 Mc/sec away from that of $K1$. A tuned intermediate-frequency (IF) amplifier with a band width of about 2.5 Mc/sec then amplifies the resulting 30-Mc/sec signal. The output of this amplifier is detected in a diode second detector, and then fed through a low-pass filter to an oscilloscope. Alternatively, the output from the second detector may be sent to a 33-cps lock-in detector which is synchronized with mechanical beam chopper inside the vacuum system.

Most of the measurements were taken without the lock-in circuit. Klystrons $K1$ and $K2$ were frequency-modulated by applying to their repellers a sawtooth sweep voltage at a rate of about 20 cps from an oscilloscope. The output of the diode second detector was then fed through a low pass filter with a bandpass of about 200 cps, and displayed on the scope. The emission lines appeared on the scope, as in Fig. 4 when the klystron $K1$ swept through the ammonia transition frequencies.

The line frequencies were measured in the usual way by means of standard frequency markers at 30-Mc/sec intervals generated by multiplication up from a 50-kc/sec quartz crystal oscillator whose frequency is monitored directly against radio station WWV.[12] An AM receiver tuned to the frequency difference between the ammonia line frequency and the nearest standard marker picks up the beat note between the signal klystron and the standard marker. As the klystron sweeps, a pip emerges from the receiver which is displayed on the oscilloscope along with the emission line pattern. Calibration of the receiver setting gives the line frequency when the pip is centered on the line.

The lock-in detector was also used during a few of the frequency measurements. In this case the difference between the signal klystron frequency and a nearby standard frequency was fed into an FM reciver, and the discriminator output of the receiver was used as an *AFC* to fix the frequency of the signal klystron at a point determined by the receiver tuning. Point by point measurements at intervals of a few kilocycles were then made in order to take advantage of the long time-constants (about $\frac{1}{3}$ sec) afforded by the lock-in circuit. The increase in sensitivity afforded by the reduced band width of the lock-in system was, however, offset by a decrease in resolution caused by the introduction of spurious frequency modulation into the klystron signal from noise in the frequency standard. In addition this method was considerably more cumbersome and time-consuming and hence was not used very much.

The equivalent band with for a detecting system of the type used has been discussed by Geschwind,[13] who showed that it is

$$\Delta\nu=[(\Delta\nu_1)(\Delta\nu_2)]^{\frac{1}{2}},$$

where $\Delta\nu_1=2.5\times10^6$ cps is the band width of the IF amplifier, and $\Delta\nu_2$ is the band width of the output circuit of the diode second detector. In the first and most frequently used measuring technique discussed above, $\Delta\nu_2$ was about 200 cps, so that

$$\Delta\nu=(2.5\times10^6\times2\times10^2)^{\frac{1}{2}}=2.2\times10^4 \text{ cps}.$$

An increased signal-to-noise ratio is therefore possible if the band width of the detecting system is reduced.

[12] Gordy, Smith, and Trambarulo, *Microwave Spectroscopy* (John Wiley and Sons, Inc., New York, 1953).

[13] S. Geschwind, thesis, Columbia University, 1951 (unpublished). See also Ann. N. Y. Acad. Sci. **55**, 751 (1952), Article 5.

The measurements were made difficult by the width of the frequency standard marker pips. This width is caused by spurious frequency modulation of the standard and klystron signals. At best, the marker pips could be narrowed to about 10 or 15 kc/sec, and the center of the pips could only be determined to about 1 kc/sec. The situation, plus the fact that drifts in the klystron control circuits made it difficult to keep the lines on the scope for very long, made the measurements accurate only to about one kilocycle, or about one seventh of the line width. In addition, most of the lines measured were unresolved structures whose shapes could not be accurately calculated because of the effects of the focuser on the populations of the various hyperfine levels. Thus, while measurements could be repeated to an accuracy of about 1 kc/sec, errors in the determinations of the positions of incompletely resolved lines were perhaps somewhat greater. The theoretical frequencies listed in Table II represent our best estimate of the frequencies determined from the theoretical patterns, which should correspond to the measured frequencies.

THE MEASUREMENTS

Measurements were taken on the hyperfine structure of the 1–1, 2–2, 3–3, and 4–3 lines of the NH_3 inversion spectrum. These measurements allowed accurate experimental values to be obtained for the constants A and C of the hydrogen magnetic interaction [see Eq. (5)]. They also provided accurate checks on the previously determined magnetic constants B, a, and b, and of the quadrupole coupling constant eqQ and its dependence on J and K. A slight asymmetry of the hyperfine structure of the 3–3 line was observed and explained by assuming that the magnitude of eqQ is 4 ± 1 kc/sec larger in the lower inversion state than in the upper; however, an apparent slight asymmetry in the structure of the 4–3 line is still unexplained.

Best values for the magnetic constants are, in kc/sec,

$$A=-17.3\pm0.5,\quad B=-14.1\pm0.3,\quad C=-2.0\pm1.0,$$
$$a=6.66\pm0.2,\quad b=6.66\pm0.2,$$

and the quadrupole coupling constant is given by

$$\langle eqQ\rangle_{Av}=4089\{1+7.7\times10^{-5}[J(J+1)+K^2]\}\pm1.5 \text{ kc/sec},$$

where $\langle eqQ\rangle_{Av}$ is the average of the quadrupole coupling constants for the upper and lower inversion levels. This result was chosen to provide a best fit to both our data and those of GM I. Our own data on the relatively low J and K lines are somewhat better fitted by a formula

$$\langle eqQ\rangle_{Av}=-4092.4[1+5\times10^{-5}J(J+1)]\pm1.5 \text{ kc/sec},$$

which, however, does not agree well at high J and K with the results of GM I.

The magnetic structures of the inversion lines may be divided into two classes; those for which $K=1$, and those for which $K\neq1$. For the latter case the K degeneracy (this is discussed in some detail in GM II) is not lifted by the hyperfine interactions. Hence the hyperfine structure of the upper and lower inversion levels is almost identical, and so the hyperfine structure of lines with $K\neq1$ is almost symmetrical about the position of a line undisplaced by hyperfine structure. The only departure from symmetry for which we have allowed is the difference in the quadrupole coupling constants for the upper and lower inversion states.

FIG. 4. Composite oscilloscope photograph of the $N^{14}H_3$ $J=K=3$ inversion transition at 23 870 Mc/sec. Below: main line with satellites caused by reorientation of the hydrogen spins. Above left: Inner quadrupole satellite, showing structure due to the hydrogens. Above right: outer quadrupole satellite. The main line has been retouched to make it more visible in the reproduction.

Figures 4 and 5 show the experimental and theoretical structures, respectively, of the 3–3 line. This is the most intense of the observed lines and is an example of the structure to be expected for the $K\neq1$ lines. The main line, composed of the twelve individual hyperfine transitions for which $\Delta F=\Delta F_1=0$, is closely flanked on either side by the magnetic satellites. The satellite on the high-frequency side of the main line is composed of the nine transitions for which $\Delta F=+1$, $\Delta F_1=0$, while the one on the low frequency side has $\Delta F=-1$, $\Delta F_1=0$. At frequency intervals of 1.7 and 2.3 Mc/sec on either side of the main line lie the quadrupole satellites, each of which as its own magnetic structure. This magnetic hyperfine structure of both the main line and the quadrupole satellites had not previously been observed due to lack of resolution or sensitivity. Because of the adiabatic focusing, the quadrupole satellites on the low-frequency side of the main line are considerably weaker than those on the high-frequency side, and only the upper two are shown in Fig. 4. It is to be noticed that the slight asymmetry of the magnetic satellites of the main line is well accounted for by the assumed variation of the quadrupole coupling constant.

Figure 6 shows the theoretical pattern for the 1–1 line. Because of the removal of the K degeneracy by the magnetic interactions in this case, the symmetry of the structure has disappeared. This line was weaker than any of the others, so somewhat larger experimental errors were incurred. It is the only one of the $K=1$ lines which was measured.

TABLE II. Hyperfine spectrum of NH_3. Frequencies are given in kc/sec relative to a theoretical line undisplaced by hyperfine structure. The theoretical frequencies listed are best estimates of the measured positions, after allowances for the observed line shapes and the width of the frequency standard pips.

Component	Transitions $F,F-F',F'$	Experimental frequency	Theoretical frequency
$J=1,\ K=1;\ A+\frac{1}{2}C=-18.3\pm0.3$ kc/sec, $B=-14.1\pm0.3$ kc/sec			
Main line	$2, 5/2 \rightarrow 2, 5/2$		
	$1, \frac{3}{2} \rightarrow 1, \frac{3}{2}$	-15.6	-14.8
	$2, \frac{3}{2} \rightarrow 2, 5/2$		
	$1, \frac{3}{2} \rightarrow 1, \frac{1}{2}$		
Inner quadrupole satellites	$2, \frac{3}{2} \rightarrow 2, \frac{3}{2}$		
	$2, 5/2 \rightarrow 2, \frac{3}{2}$	13.7	13.0
	$1, \frac{3}{2} \rightarrow 1, \frac{3}{2}$		
	$1, \frac{1}{2} \rightarrow 1, \frac{1}{2}$		
	$2, \frac{3}{2} \rightarrow 2, \frac{1}{2}$	-624.2	-622.9
	$2, 5/2 \rightarrow 1, \frac{3}{2}$	-591.3	-590.6
	$2, \frac{3}{2} \rightarrow 1, \frac{1}{2}$		
	$1, \frac{1}{2} \rightarrow 2, \frac{1}{2}$	575.2	572.6
	$1, \frac{1}{2} \rightarrow 2, \frac{3}{2}$		
	$1, \frac{1}{2} \rightarrow 2, 5/2$	619.2	617.8
Outer quadrupole satellites	$0, \frac{1}{2} \rightarrow 1, \frac{1}{2}$		1569.0
	$0, \frac{1}{2} \rightarrow 1, \frac{3}{2}$		1526.9
	$1, \frac{3}{2} \rightarrow 0, \frac{1}{2}$	1540.3	1540.8
	$1, \frac{1}{2} \rightarrow 0, \frac{1}{2}$		

Component	Transitions	No. individual contributing lines	Experimented frequency	Theoretical frequency
$J=2,\ K=2;\ A+\frac{1}{2}C=-18.6\pm0.2$ kc/sec				
Main line	$\Delta F_1=\Delta F=0$	6	$+0.3$	-0.3
Magnetic satellites	$\Delta F_1=0,\ \Delta F=-1$	3	-42.2	-42.0
	$\Delta F_1=0,\ \Delta F=+1$	3	41.1	40.7
Inner quadrupole satellites	$F_1=3\rightarrow2,\ \Delta F=-1$	2	-1297.3	-1297.3
	$F_1=3\rightarrow2,\ \Delta F=0$	1		-1255
	$F_1=2\rightarrow3,\ \Delta F=0$	1	1255.2	1256.0
	$F_1=2\rightarrow3,\ \Delta F=+1$	2	1296.5	1296.6
Outer quadrupole satellites	$F_1=1\rightarrow2,\ \Delta F=0$	1		-2099.8
	$F_1=1\rightarrow2,\ \Delta F=+1$	2	-2056.4	-2057.5
	$F_1=2\rightarrow-1,\ \Delta F=-1$	2	2056.8	2057.5
	$F_1=2\rightarrow1,\ \Delta F=0$	1	2100.6	2099.8

Component	Transitions	No. individual contributing lines	Experimented frequency	Theoretical frequency
$J=3,\ K=3;\ A+\frac{1}{2}C=-18.8\pm0.2$ kc/sec				
Main line	$\Delta F_1=\Delta F=0$	12		-0.5
Magnetic satellites	$\Delta F_1=0,\ \Delta F-1$	9	-71.0	-71.0
			-62.5	-62.5
	$\Delta F_1=0,\ \Delta F=+1$	9	60.9	61.0
			70.4	70.3
Inner quadrupole satellites	$F_1=4\rightarrow3,\ \Delta F=0$	3		
	$F_1=4\rightarrow3,\ \Delta F=-1$	4		-1689.4
				-1680.0
	$F_1=3\rightarrow4,\ \Delta F=+1$	4	1680	1679.2
			1689.0	1688.6
	$F_1=3\rightarrow4,\ \Delta F=0$	3		
Outer quadrupole satellites	$F_1=2\rightarrow3,\ \Delta F=0$	3		-2365.3
				-2383.8
	$F_1=2\rightarrow3,\ \Delta F=+1$	4		-2323.5
				-2311.2
				-2302.1
	$F_1=3\rightarrow2,\ \Delta F=-1$	4	2303.0	2302.1
			2311.0	2311.2
			2323.0	2323.5
	$F_1=3\rightarrow2,\ \Delta F=0$	3	2365.0	2364.8
			2383.5	2383.0
$J=4,\ K=3;\ A+0.45C=-18.2\pm0.2$ kc/sec				
Main line	$\Delta F=\Delta F_1=0$	12	-1.7	0
Magnetic satellites	$\Delta F_1=0,\ \Delta F=-1$	9	-82.0	-81.5
			-65.5	-65.5
	$\Delta F_1=0,\ \Delta F=+1$	9	64.8	65.5
			81.6	81.5
Inner quadrupole satellites	$F_1=5\rightarrow4,\ \Delta F=-1$	3	-460.9	-462.4
	$F_1=4\rightarrow5,\ \Delta F=+1$	3	-464.2	462.6
Outer quadrupole satellites	$F_1=3\rightarrow4,\ \Delta F=+1$	3	-638.4	638.0
	$F_1=4\rightarrow3,\ \Delta F=-1$	3	638.8	638.2

Table II is a list of all the experimentally measured frequencies, along with the corresponding frequencies determined from theoretical plots of the line structures. In most cases, the observed structures were the envelopes of groups of unresolved or partially resolved individual lines. If these structures had several resolved peaks, then wherever possible the positions of the various peaks were measured. In cases where the envelope was a weak unresolved asymmetric structure, the measurement usually corresponded more nearly to the midpoint between the two frequencies at which the emission was at half-maximum rather than to the point of maximum emission. This was so since the width of the frequency standard pip used for measuring was as broad or broader than the line structures involved, and such a piper could be much more easily centered between the steeply sloping sides of a peak than on its relatively indistinct maximum. In some cases errors greater than one kc/sec could be expected, since effects of the adiabatic focusing on the populations of the various hyperfine levels could not be exactly calculated, therefore the calculation of the line shapes for the unresolved structures was subject to some error.

From the spacings of the magnetic satellites of each of the three measured $K\neq1$ lines, values of the quantity $\{A+CK^2/[J(J+1)]\}$ could be obtained to an accuracy of about 0.2 kc/sec. These values are shown in Table II. From them the values of A and C given above were calculated. A best fit was then found for the 1–1 line using these values of A and C and varying B. The value of B so obtained is also given above. Theoretically $\{A+CK^2/[J(J+1)]\}$ and B could both be obtained

from the structure of this line, but since its intensity was small and the errors correspondingly large, it seemed better to determine A and C from the other lines.

Constants for the magnetic energy associated with orientation of the hydrogen nuclei were used to calculate the structures of the quadrupole satellites, and to obtain the positions of lines displaced only by the hyperfine interactions of the nitrogen nucleus. These displacements were then used to calculate values of $\langle eqQ\rangle_{hv}$ and of $\{a+(b+a)K^2/[J(J+1)]\}$, the quadrupole and magnetic dipole coupling constants. The second expression given above for the value of $\langle eqQ\rangle_{hv}$ as a function of J and K seemed to fit our data best, and it was used to calculate the theoretical line frequencies listed in Table II. The first expression for $\langle eqQ\rangle_{hv}$ agrees better with the results of GM I at high values of J and K, but it increases some of the discrepancies between our own theoretical and experimental results for the case of the 1–1 line. It was, however, thought to be the best available compromise between the present work and that of GM I. The form of the dependence of $\langle eqQ\rangle_{hv}$ on J and K given in GM I was used simply because this form could not be well determined from the few lines which were measured under high resolution.

These results give a considerable amount of information about the electric and magnetic fields existing within the molecule, which should be explainable in terms of a satisfactory molecular theory. All three components of the molecular magnetic field at the position of the hydrogen nuclei are now known to good accuracy, as well as the two different components of the magnetic field and the gradient of the electric field at the position of the nitrogen nucleus, (which had been previously determined with good accuracy). Some discussion of the interpretation of these constants has been made in GM I.

One expects small variations of eqQ with J and K on account of the centrifugal distortion of the molecule, which produces changes in the inversion wave function. Similarly, a slight change in eqQ from upper inversion level to lower inversion level can be expected. The change of 4 kc/sec which explained satisfactorily the asymmetry of the 3–3 line was also applied on calculating the hyperfine structures of the other lines, but produced no changes in the spectra which would be detectable with the present apparatus. A remaining discrepancy which has not yet been explained is the apparent displacement of the main line of the 4–3 transition. Table II shows that the main line appears to be displaced by about 3 kc/sec away from the average frequency of the two inner quadrupole satellites. Since these measurements were taken rather carefully, this shift appears to be outside of experimental error. In this line the quadrupole splittings are small, so that the asymmetry cannot be accounted for by the small change in quadrupole coupling constant.

The change in eqQ from the upper state to lower can be qualitatively discussed with a relatively crude model. In the lower inversion state, the inversion wave function is a symmetrical function of z, the coordinate which

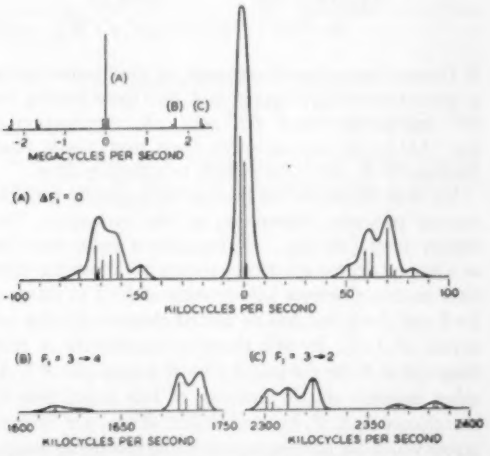


FIG. 5. Theoretical pattern of the hyperfine structure of the $J=K=3$ inversion transition. $\mathbf{F}_1=\mathbf{J}+\mathbf{I}_N$, where I_N is the spin of the nitrogen nucleus. The box at the upper left shows the complete spectrum. Below, the components (a), (b), and (c) of the complete spectrum are shown expanded. Frequency displacements from the position of a line undisplaced by hyperfine structure are shown on the figure. The positions and heights of the solid lines indicate the frequencies and intensities of the various hyperfine structure components of this transition. A total width at half-maximum of 7 kc/sec for each individual component was chosen to give a good fit to the observed spectrum. The intensity of the main line has been reduced by a factor of five.

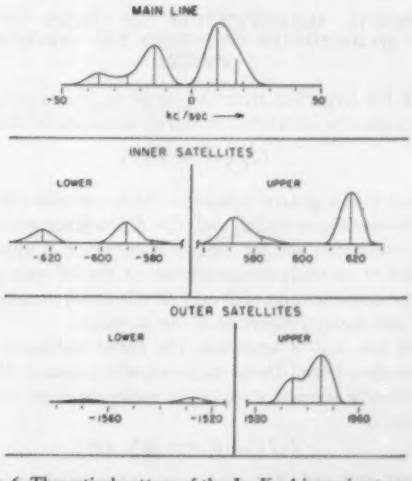


FIG. 6. Theoretical pattern of the $J=K=1$ inversion transition. Note that the symmetry of the hyperfine structure, which was evident in the 3,3 transition, is not present in this case. The total width at half-maximum of each hyperfine component of the transition is chosen as 10 kc/sec. Frequency displacements from the position of a line undisplaced by hyperfine structure are shown on the figure.

gives the position of the nitrogen nucleus on the symmetry axis of the molecule. In the upper state, the inversion wave function is an odd function of z. The nitrogen thus spends more of its time near the plane of the three hydrogens when the molecule is in the lower state. In its most probable configuration, the bond angles of the molecule are only slightly less than the tetrahedral angle. The quadrupole coupling constant is considerably smaller than would be expected if the five bonding electrons combined to form a completely covalent tetrahedral bond. This is due to the partially ionic character of the bond.[14] If, however, we neglect the ionic character and assume a completely covalent bond, we shall observe the proper trend in eqQ. For the tetrahedral bond, the two nonbonding electrons would be represented by $\frac{1}{2}s+\frac{1}{2}\sqrt{3}p_z$ hybrid orbitals, where z is the molecular axis. The value of $q_z=\partial^2V/\partial z^2$ at the position of the nitrogen may be considered as produced by only one of these nonbonding electrons, since the other plus the three bonding electrons form a symmetrical structure for which $q_z=0$. This results in a value $q_z=\frac{3}{4}q_p$, where q_p is the value of $\partial^2V/\partial z^2$ produced by a pure p_z orbital. For the planar configuration of the molecule, through which it must go while inverting, the three bonding electrons should be in sp^2 hybrid orbitals, while the two nonbonding electrons would both be in pure p_z orbitals. It can be shown that in this configuration $q_z=q_p$, which is larger in magnitude than that given by the tetrahedral structure. Thus one should expect the observed larger magnitude for quadrupole coupling in the symmetric lower inversion state, in which the nitrogen spends more time near the plane of the hydrogens.

APPENDIX. SIMPLIFICATION OF TERMS OF THE HAMILTONIAN INVOLVING THE OFF-AXIS NUCLEI

In the hyperfine structure problem in ammonia, one is frequently interested in matrix elements of the form

$$\langle \gamma JK|O|\gamma' J'K'\rangle, \tag{A1}$$

where O is a general operator which contains the spins of the hydrogen nuclei, and $J_z=K$. In other symmetric top molecules similar matrix elements will appear in which O contains the properties of the off axis nuclei. Some simplification of O may be effected by considering the symmetry properties of the molecule.

In the case of ammonia, the three hydrogen nuclei must obey Fermi-Dirac statistics, which means that any physically allowed state of the molecule must obey the equation

$$P_{ij}^{\sigma}P_{ij}^{\sigma}|\psi\rangle=-|\psi\rangle \quad i\neq j, \tag{A2}$$

where P_{ij}^{σ} and P_{ij}^{r} represent exchange of space and spin coordinates, respectively, of hydrogen nuclei i and j. We shall find it to our advantage to confine our attention to those states of the molecule which satisfy the less restrictive equation

$$P_{ijk}^{\sigma}P_{ijk}^{x}|\psi\rangle=|\psi\rangle \quad i\neq j\neq k, \tag{A3}$$

where $P_{ijk}=P_{ij}P_{jk}$. States can be constructed satisfying Eq. (A3) which are separable into the product of a spin function and a space function and can be treated without difficulty.[15] The space functions are to be eigenfunctions of the angular momentum operator J_z, which commutes with both P_{ijk}^{x} and P_{ijk}^{σ}. The final state function satisfying Eq. (A2) will of necessity be formed of linear combinations of functions satisfying Eq. (A3), so that no generality is lost in this procedure.

The operation P_{123}^{x} is equivalent to a rotation of the molecule through an angle of $2\pi/3$ around its axis of symmetry, and thus can be written in the form $P_{123}^{x}=e^{\frac{1}{3}\pi iJ_z}$. Also P_{123}^{σ} can be expressed in terms of the hydrogen spins I_i.[16] It has the effect of interchanging hydrogen spins; i.e., $(P_{123}^{\sigma})^{-1}\mathbf{I}_2P_{123}^{\sigma}=\mathbf{I}_1$, etc. In view of Eq. (A3) we may write

$$\begin{aligned}&\langle \gamma JK|O|\gamma' J'K'\rangle\\&=\langle \gamma JK|(P_{123}^{x})(P_{123}^{\sigma})^{-1}OP_{123}^{\sigma}P_{123}^{x}|\gamma' J'K'\rangle\\&=\langle \gamma JK|(P_{132}^{x})^{-1}(P_{132}^{\sigma})^{-1}OP_{132}^{\sigma}P_{132}^{x}|\gamma' J'K'\rangle.\end{aligned} \tag{A4}$$

Allowing only the P^{x} operators to operate on the state functions, and averaging the three resulting expressions gives

$$\begin{aligned}&\langle \gamma JK|O|\gamma' J'K'\rangle\\&\quad=\tfrac{1}{3}\langle \gamma JK|O+e^{\frac{1}{3}\pi i(K'-K)}P_{132}^{\sigma}OP_{123}^{\sigma}\\&\qquad+e^{-\frac{1}{3}\pi i(K'-K)}P_{123}^{\sigma}OP_{132}^{\sigma}|\gamma' J'K'\rangle.\end{aligned} \tag{A5}$$

In the particular case of ammonia we shall be interested in matrix elements diagonal in K and those joining the two degenerate states $+K$ and $-K$. Application of Eq. (A5) to the various interactions, particularly those diagonal in K, will reduce them to a simpler form.

We shall illustrate the method by applying it to the mutual spin-spin interaction of the hydrogens. This energy term [see Eq. (5)] transforms under rotations as a second degree spherical harmonic.[17] It can therefore have matrix elements joining state of $I=\frac{3}{2}$ to states of $I=\frac{3}{2}$ and $I=\frac{1}{2}$, but has no matrix elements joining two states of $I=\frac{1}{2}$. It will therefore contribute a term diagonal in K for the states with K a multiple of 3. All other nonzero matrix elements of this interaction (it can change J, K, F_1 by a maximum of two) are between states which are nondegenerate in the rotational energy, and so these matrix elements can be ignored. We are interested therefore only in matrix elements diagonal

[14] J. Bardeen and C. H. Townes, Phys. Rev. 73, 97 (1948).

[15] For example, P_{123}^{x} operating on the rotation vibration function $\psi_v{}^{JKM}$ multiplies it by $e^{\frac{2}{3}\pi iK/3}$. Thus we choose a spin function with the property that P_{123}^{σ} operating on its multiplies it by $e^{-\frac{2}{3}\pi iK/3}$.

[16] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, 1947), third edition, p. 222.

[17] J. H. Van Vleck, *Theory of Electric and Magnetic Susceptibilities* (Oxford University Press, London, 1932).

in K. Substituting $K'=K$ in Eq. (A5) gives

$$O \rightarrow \tfrac{1}{4}\{O+P_{123}{}^{\sigma}OP_{123}{}^{\sigma}+P_{123}{}^{\sigma}OP_{132}{}^{\sigma}\}. \quad \text{(A6)}$$

This operation immediately reduces (1) to (2) with only a small amount of algebra, after (1) is expanded in terms of the molecular axes.

To demonstrate the application of the method to interactions which have matrix elements of interest which are off-diagonal in K, we shall apply it to the hydrogen-nitrogen spin-spin interaction. This part of the Hamiltonian may be written

$$G^{\mathrm{HN}}=g_{\mathrm{H}}g_{\mathrm{N}}\mu_0{}^2r_{\mathrm{HH}}{}^{-3} \times\sum_i[\mathbf{I}_i\cdot\mathbf{I}_{\mathrm{N}}-3r_{\mathrm{HN}}{}^{-2}(\mathbf{I}_i\cdot\mathbf{r}_{i\mathrm{N}})(\mathbf{I}_{\mathrm{N}}\cdot\mathbf{r}_{i\mathrm{N}}), \quad \text{(A7)}$$

where r_{HN} is the hydrogen-nitrogen distance, and $\mathbf{r}_{i\mathrm{N}}$ is the vector distance between the ith hydrogen and the nitrogen. This interaction can lift the K degeneracy for $K=\pm1$, as shown in GM II, and so we shall be interested in the diagonal elements and those joining $K=\pm1$ to $K=\mp1$. Again, all other nonzero matrix elements are off-diagonal in the rotational energy. For the diagonal part we may make the substitution (A6), which here has the effect of replacing I_i by $I/3$, and so we obtain

$$\langle K|G_1{}^{\mathrm{HN}}|K\rangle=g_{\mathrm{H}}g_{\mathrm{N}}\mu_0{}^2\langle K|r_{\mathrm{HN}}{}^{-3}[\mathbf{I}\cdot\mathbf{I}_{\mathrm{N}} -\textstyle\sum_i r_{\mathrm{HN}}{}^{-2}(\mathbf{I}\cdot\mathbf{r}_{i\mathrm{N}})(\mathbf{I}_{\mathrm{N}}\cdot\mathbf{r}_{i\mathrm{N}})]|K\rangle. \quad \text{(A8)}$$

Expanding this in terms of the molecular axes then results in

$$\langle K|G_1{}^{\mathrm{HN}}|K\rangle=2g_{\mathrm{H}}g_{\mathrm{N}}v_0{}^2\langle K|r_{\mathrm{HN}}{}^{-3}(1-\tfrac{3}{2}\sin^2\beta) \times(I\cdot I_{\mathrm{N}}-3I_aI_{\mathrm{N}a})|K\rangle. \quad \text{(A9)}$$

Now we are here interested also in those matrix elements which join the two degenerate states of $K=\pm1$, so we must evaluate $\langle K=\pm1|G_1{}^{\mathrm{HN}}|K'=\mp1\rangle$. Application of (7) reduces this matrix to the form

$$\langle K=\pm1|G_1{}^{\mathrm{HN}}|K'=\mp1\rangle =g_{\mathrm{H}}g_{\mathrm{N}}\mu_0{}^2\langle K=\pm1|r_{\mathrm{HN}}{}^{-3}[-(9/4)\sin^2\beta \times(I_{1x}{}'\mp iI_{1y}{}')(I_{\mathrm{N}x}{}'\mp iI_{\mathrm{N}y}{}')]|K'=\mp1\rangle, \quad \text{(A10)}$$

or

$$\langle K=\pm1|G_1{}^{\mathrm{HN}}|K'=\mp1\rangle =g_{\mathrm{H}}g_{\mathrm{N}}\mu_0{}^2\langle K=\pm1|r_{\mathrm{HN}}{}^{-3}(9/2)\sin^2\beta \times(I_{1y}I_{\mathrm{N}y}-I_{1x}I_{\mathrm{N}x})|K=\mp1\rangle. \quad \text{(A11)}$$

Since (A9) and (A11) give all the matrix elements of interest, we see that the complete interaction can be replaced by the somewhat simplified form:

$$G_1{}^{\mathrm{HN}}\rightarrow g_{\mathrm{H}}g_{\mathrm{N}}\mu_0{}^2r_{\mathrm{HN}}{}^{-3}[(1-\tfrac{3}{2}\sin^2\beta)(\mathbf{I}\cdot\mathbf{I}_{\mathrm{N}}-3I_aI_{\mathrm{N}a}) +(9/2)\sin^2\beta(I_{1y}I_{\mathrm{N}y}-I_{1x}I_{\mathrm{N}x})]. \quad \text{(A12)}$$

Evaluation of the off-diagonal matrix elements (A11) may be made according to the standard methods of Condon and Shortley, remembering our choice of spin functions, or the equivalent Hamiltonian (A12) may be treated by the methods of GM II.

The method may easily be extended to the treatment of the $I\cdot J$ interaction, and yields the same result as that obtained in GM II. (Note the correction to GM II given in reference 8.)

ACKNOWLEDGMENTS

The author would like to express his appreciation to the personnel of the Columbia Radiation Laboratory who assisted in the construction of the experimental apparatus. Dr. H. J. Zeiger and Dr. T. C. Wang contributed greatly to the experiment in various stages. In particular the author is indebted to Professor C. H. Townes for suggesting this research and for his continued advice and guidance.

The Maser—New Type of Microwave Amplifier, Frequency Standard, and Spectrometer*†

J. P. GORDON,‡ H. J. ZEIGER,§ AND C. H. TOWNES
Columbia University, New York, New York

(Received May 4, 1955)

A type of device is described which can be used as a microwave amplifier, spectrometer, or oscillator. Experimental results are given. When operated as a spectrometer, the device has good sensitivity, and, by eliminating the usual Doppler broadening, a resolution of 7 kc/sec has been achieved. Operated as an oscillator, the device produced a frequency stable to at least 4 parts in 10^{12} in times of the order of a second, and stable over periods of an hour or more to at least a part in 10^{10}. The device is examined theoretically, and results are given for the expected sensitivity of the spectrometer, the stability and purity of the oscillation, and the noise figure of the amplifier. Under certain conditions a noise figure approaching the theoretical limit of unity, along with reasonably high gain, should be attainable.

INTRODUCTION

A TYPE of device is described below can be used as a microwave spectrometer, a microwave amplifier, or as an oscillator. As a spectrometer, it has good sensitivity and very high resolution since it can virtually eliminate the Doppler effect. As an amplifier of microwaves, it should have a narrow band width, a very low noise figure and the general properties of a feedback amplifier which can produce sustained oscillations. Power output of the amplifier or oscillator is small, but sufficiently large for many purposes.

The device utilizes a molecular beam in which molecules in the excited state of a microwave transition are selected. Interaction between these excited molecules and a microwave field produces additional radiation and hence amplification by stimulated emission. We call an apparatus utilizing this technique a "maser," which is an acronym for "microwave amplification by stimulated emission of radiation."

Some results obtained with this device have already been briefly reported.[1] An independent proposal for a system of this general type has also been published.[2] We shall here examine in some detail the general behavior and characteristics of the maser and compare experimental results with theoretical expectations. Particular attention is given to its operation with ammonia molecules. The preceding paper,[3] which will hereafter be referred to as (I), discusses an investigation of the hyperfine structure of the microwave spectrum of $N^{14}H_3$ with this apparatus. Certain of its properties which are necessary for an understanding of the relative intensities of the hyperfine structure components are also discussed there.

BRIEF DESCRIPTION OF OPERATION

A molecular beam of ammonia is produced by allowing ammonia molecules to diffuse out a directional source consisting of many fine tubes. The beam then transverses a region in which a highly nonuniform electrostatic field forms a selective lens, focusing those molecules which are in upper inversion states while defocusing those in lower inversion states. The upper inversion state molecules emerge from the focusing field and enter a resonant cavity in which downward transitions to the lower inversion states are induced. A simplified block diagram of this apparatus is given in Fig. 1. The source, focuser, and resonant cavity are all enclosed in a vacuum chamber.

For operation of the maser as a spectrometer, power of varying frequency is introduced into the cavity from an external source. The molecular resonances are then observed as sharp increases in the power level in the cavity when the external oscillator frequency passes the molecular resonance frequencies.

At the frequencies of the molecular transitions, the beam amplifies the power input to the cavity. Thus the maser may be used as a narrow-band amplifier. Since the molecules are uncharged, the usual shot noise existing in an electronic amplifier is missing, and essentially no noise in addition to fundamental thermal noise is present in the amplifier.

If the number of molecules in the beam is increased beyond a certain critical value the maser oscillates. At the critical beam strength a high microwave energy density can be maintained in the cavity by the beam alone since the power emitted from the beam compensates for the power lost to the cavity walls and coupled wave guides. This oscillation is shown both experimentally and theoretically to be extremely monochromatic.

* Work supported jointly by the Signal Corps, the Office of Naval Research, and the Air Research and Development Command.

† Submitted by J. P. Gordon in partial fulfillment of the requirements of the degree of Doctor of Philosophy at Columbia University.

‡ Now at the Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

§ Carbide and Carbon Postdoctoral Fellow in Physics, now at Project Lincoln, Massachusetts Institute of Technology, Cambridge, Massachusetts.

[1] Gordon, Zeiger, and Townes, Phys. Rev. **95**, 282 (1954).

[2] N. G. Bassov and A. M. Prokhorov, J. Exptl. Theoret. Phys. (U.S.S.R.) **27**, 431 (1954). Also N. G. Bassov and A. M. Prokhorov, Proc. Acad. of Sciences (U.S.S.R.) **101**, 47 (1945).

[3] J. P. Gordon, preceding paper [Phys. Rev. **99**, 1253 (1955)].

APPARATUS

The geometrical details of the apparatus are not at all critical, and so only a brief description of them will be made. Two ammonia masers have been constructed with somewhat different focusers. Both have operated satisfactorily.

A source designed to create a directional beam of the ammonia molecules was used. An array of fine tubes is produced in accordance with a technique described by Zacharias,[4] which is as follows. A $\frac{1}{4}$ in. wide strip of 0.001-in. metal foil (stainless steel or nickel, for example) is corregated by rolling it between two fine-toothed gears. This strip is laid beside a similar uncorregated strip. The corregations then form channels leading from one edge of the pair of strips to the other. Many such pairs can then be stacked together to create a two-dimensional array of channels, or, as was done in this work, one pair of strips can be rolled up on a thin spindle. The channels so produced were about 0.002 in. by 0.006 in. in cross section. The area covered by the array of channels was a circle of radius about 0.2 in., which was about equal to the opening into the focuser. Gas from a tank of anhydrous ammonia was maintained behind this source at a pressure of a few millimeters of mercury.

This type of source should produce a strong but directed beam of molecules flowing in the direction of the channels. It proved experimentally to be several times more effective than a source consisting of one annular ring a few mils wide at a radius of 0.12 in., which was also tried.

The electrodes of the focuser were arranged as shown in Fig. 1. High voltage is applied to the two electrodes marked V, while the other two are kept at ground. Paul *et al.*[5,6] have used similar magnetic pole arrangements for the focusing of atomic beams.

In the first maser which was constructed the inner faces of the electrodes were shaped to form hyperbolas with 0.4-in. separating opposing electrodes. The distance of closest approach between adjacent electrodes was 0.08 in., and the focuser was about 22 in. long. Voltages up to 15 kv could be applied to these electrodes before sparking occurred. In the second maser the electrodes were shaped in the same way, but were separated from each other by 0.16 in. This allowed voltages up to almost 30 kv to be applied, and somewhat more satisfactory operation was obtained since higher field gradients could be achieved in the region between the electrodes. This second focuser was only 8 in. long. Teflon spacers were used to keep the electrodes in place. To provide more adequate pumping of the large amount of ammonia released into the vacuum system from the source the focuser electrodes were hollow and were filled with liquid nitrogen.

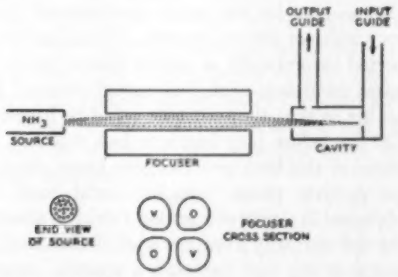


FIG. 1. Simplified diagram of the essential parts of the maser.

The resonant cavities used in most of this work were circular in cross section, about 0.6 in. in diameter by 4.5 in. long, and were resonant in the TE_{011} mode at the frequency of interest (about 24 kMc/sec). Each cavity could be turned over a range of about 50 Mc/sec by means of a short section of enlarged diameter and variable length at one end. A hole 0.4 in. in diameter in the other end allowed the beam to enter. The beam traversed the length of the cavity. The cavities were made long to provide a considerable time for the molecules to interact with the microwave field. Only one-half wavelength of the microwave field in the cavity in the axial direction was allowed for reasons which will appear later in the paper. Since the free space wavelength of 24-kMc/sec microwaves is only about 0.5 in., and an axial wavelength of about 9 in. was required in the cavity, the diameter of the cavity had to be very close to the cut-off diameter for the TE_{01} mode in circular wave guide. The diameter of the beam entrance hole was well beyond cutoff for this mode and so very little loss of microwave power from it was encountered. The cavities were machined and mechanically polished. They were made of copper or silver-plated Invar, and had values of Q near 12 000. Some work was also done with cavities in the TM_{01} mode which has some advantages over the TE_{01} mode. However, the measurements described here all apply to the TE_{011} cavities.

Microwave power was coupled into and out from the cavities in several ways. Some cavities had separate input and output wave guides, power being coupled into the cavity through a two-hole input in the end of the cavity furthest from the source and coupled out through a hole in the sidewall of the cavity. In other cavities the sidewall hole served as both input and output, and the end-wall coupling was eliminated. About the same spectroscopic sensitivity was obtained with both types of cavities.

Three MCF 300 diffusion pumps (Consolidated Vacuum Company, Inc.) were used to maintain the necessary vacuum of less than 10^{-5} mm Hg. Nevertheless, due to the large volume of gas released into the system through the source, satisfactory operation has

[4] J. R. Zacharias and R. D. Haun Jr., Quarterly Progress Report, Massachusetts Institute of Technology Research Laboratory of Electronics, 34, October, 1954 (unpublished).
[5] H. Friedberg and W. Paul, Naturwiss. 38, 159 (1951).
[6] H. G. Bennewitz and W. Paul, Z. Physik 139, 489 (1954).

not yet been attained without cooling the focuser electrodes with liquid nitrogen. At 78°K the vapor pressure of ammonia is considerably less than 10^{-6} mm Hg and so the cold electrode surfaces provide a large trapping area which helps maintain a sufficiently low pressure in the vacuum chamber. The pumping could undoubtedly be accomplished by liquid air traps alone; however the diffusion pumps alone have so far proven insufficient. The solidified ammonia which builds up on the focuser electrodes is somewhat of a nuisance as electrostatic charges which distort the focusing field tend to build up on it, and crystals form which can eventually impede the flow of gas. For the relatively short runs, however, which are required for spectroscopic work, this arrangement has been fairly satisfactory.

EXPERIMENTAL RESULTS

Experimental results have been obtained with the maser as a spectrometer and as an oscillator. Although it has been operated as an amplifier, there has as yet been no measurement of its characteristics in this role. Its properties as an amplifier are examined theoretically below.

The reader is referred to (I) for the results obtained from an examination of the hyperfine structure of the $N^{14}H_3$ inversion spectrum with the maser. Resolution of about seven kc/sec was obtained, which is a considerable improvement over the limit of about[7] 65 kc/sec imposed by Doppler broadening in the usual absorption-cell type of microwave spectrometer. This resolution can be improved still further by appropriate cavity design. The sensitivity of the maser was considerably better than that of other spectrometers which have had comparably high resolution.[8–10]

The factors which determine the sensitivity and resolution of the maser spectrometer are discussed in detail below, but we may make a general comment here. The sensitivity of the maser depends in part on the physical separation of quantum states by the focuser and thus on the forces exerted by the focuser on molecules in the various quantum states. For this reason its sensitivity is not simply related to the gas absorption coefficient for a given molecular transition. Each individual case must be examined in detail. Due to the focuser, for example, the sensitivity of the maser varies more rapidly with the dipole moment of the molecule to be studied than does that of the ordinary absorption spectrometer.

The experimental results obtained with the maser in its role as an oscillator agree with the theory given below and show that its oscillation is indeed extremely monochromatic, in fact more monochromatic than any other known source of waves. Oscillations have been produced at the frequencies of the 3–3 and 2–2 inversion lines of the ammonia spectrum, those for the 3–3 line being the stronger. Tests of the oscillator stability were made using the 3–3 line, so we shall limit the discussion to oscillation at this frequency. Other ammonia transitions, or transitions of other molecules could, of course, be used to operate a maser oscillator.

The frequency of the $N^{14}H_3$ 3–3 inversion transition is 23 870 mc/sec. The maser oscillation at this frequency was sufficiently stable in an experimental test so that a clean audio-frequency beat note between the two masers could be obtained. This beat note, which was typically at about 30 cycles per second, appeared on an oscilloscope as a perfect sine wave, with no random phase variations observable above the noise in the detecting system. The power emitted from the beams during this test was not measured directly, but is estimated to be about 5×10^{-10} watt.

The test of the oscillators was made by combining signals from the two maser oscillators together in a 1N26 crystal detector. A heterodyne detection scheme was used, with a 2K50 klystron as a local oscillator and a 30-Mc/sec intermediate-frequency (IF) amplifier. The amplified intermediate frequency signals from the two maser oscillators were then beat together in a diode detector, and their difference, which was then a direct beat between the two maser oscillator frequencies, displayed on an oscilloscope. The over-all band width of this detecting system was about 2×10^4 cps, and the beat note appeared on the oscilloscope with a signal to noise ratio of about 20 to 1.

It was found that the frequency of oscillation of each maser could be varied one or two kc/sec on either side of the molecular transition frequency by varying the cavity resonance frequency about the transition frequency. If the cavity was detuned too far, the oscillation ceased. The ratio of the frequency shift of the oscillation to the frequency shift of the cavity was almost exactly equal to the ratio of the frequency width of the molecular response (that is, the line width of the molecular transition as seen by the maser spectrometer) to the frequency width of the cavity mode. This behavior is to be expected theoretically as will be shown below. The two maser oscillators were well enough isolated from one another so that the beat note could be lowered to about 20 cps before they began to lock together. The appearance of this beat note has been noted above. As perhaps $\frac{1}{10}$-cycle phase variation could have been easily detected in a time of a second (which is about the time the eye normally averages what it observes), the appearance of the beat indicates a spectral purity of each oscillator of at least 0.1 part in 2.4×10^{10}, or 4 parts in 10^{12} in a time of the order of a second.

By using Invar cavities maintained in contact with ice water to control thermal shifts in their resonant frequencies, the oscillators were kept in operation for

[7] Gunther-Mohr, White, Schawlow, Good, and Coles, Phys. Rev. **94**, 1184 (1954).
[8] G. Newell and R. H. Dicke, Phys. Rev. **83**, 1064 (1951).
[9] R. H. Romer and R. H. Dicke, Phys. Rev. **98**, 1160(A) (1955).
[10] M. W. P. Strandberg and H. Dreicer, Phys. Rev. **94**, 1393 (1954).

periods of an hour or so with maximum variations in the beat frequency of about 5 cps or 2 parts in 10^{10} and an average variation of about one part in 10^{10}. Even these small variations seemed to be connected with temperature changes such as those associated with replenishing the liquid nitrogen supply in the focusers. Theory indicates that variations of about 0.1°C in temperature, which was about the accuracy of the temperature control, would cause frequency deviations of just this amount.

It was found that the oscillation frequency was slightly dependent on the source pressure and the focuser voltage, both of which affect the strength of the beam. These often produced frequency changes of the order of 20 cycles per second when either voltage or pressure was changed by about 25%. As the cavity was tuned, however, both these effects changed direction, and the null points for the two masers coincide to within about 30 cps. The frequency at which these effects disappear is probably very near the center frequency of the molecular response, so this may provide a very convenient way of resetting the frequency of a maser oscillator without reference to any other external standard of frequency.

THE FOCUSER

In (I) it was shown that forces are exerted by the nonuniform electric field of the focuser on the ammonia molecules, the force being radially inward toward the focuser axis for molecules in upper inversion states and radially outward for molecules in lower inversion states. Molecules in upper inversion states are therefore focused by the field, and only these molecules reach the cavity. Moreover, the quadrupole hyperfine splitting of the upper inversion state was shown to affect the focusing since the flight of the molecules through the focuser is adiabatic with respect to transitions between the different quadrupole levels. As a result the higher energy quadrupole levels are focused considerably more strongly than the lower energy ones. The further slight splitting of the various quadrupole states by the magnetic hyperfine interactions of the hydrogen nuclei has little effect since the molecules make many transitions between these closely spaced levels as they enter and leave the focuser. In regions of high field strength where hyperfine effects are unimportant and can be neglected the energy of the molecules in an electric field may be written as

$$W = W_{\text{rotation}}(J,K) \pm \left\{ \left(\frac{h\nu_0}{2} \right)^2 + \left(\frac{M_J K}{J(J+1)} \mu \mathcal{E} \right)^2 \right\}^{\frac{1}{2}}, \quad (1)$$

where ν_0 is the zero-field inversion frequency, J, K, and M_J specify the rotational state of the molecule relative to the direction of the field, μ is the molecular dipole moment, and $\mathcal{E}$ is the magnitude of the electric field.

With these considerations in mind, an approximate calculation of the total number of molecules in the upper inversion state which are trapped by the potential well of the focuser and delivered to the cavity is fairly straightforward. It involves some computation, since the line used for the oscillation (the main line of the $J=K=3$ inversion transition) is composed of three different but unresolved component transitions between quadrupole sublevels of the inversion states, and therefore the number of molecules trapped by the focuser must be calculated for each of these three sublevels and the results added. This calculation is outlined below. We shall consider in detail the properties of the first maser oscillator, with which the work reported in (I) was done.

The focuser electrodes form approximate equipotentials of the potential $V = V_0 r^2 \cos 2\theta$, where r and θ are cylindrical coordinates of a system whose z axis coincides with the axis of the focuser. 15 kv applied to the high-voltage focuser electrodes establishes an electric field whose magnitude is given by

$$\mathcal{E} = 200r, \quad (2)$$

where $\mathcal{E}$ is measured in esu and r is in cm. For simplicity we shall assume that the source is small in area and is located on the axis of the focuser. We shall also assume that all molecules which can travel farther than 0.5 cm from the focuser axis collide with the focuser electrodes and are removed from the beam. From (1) and (2) it is seen that the force ($\mathbf{f} = -\text{grad}W$) on the molecules is radial, and for small field strength is proportional to r. Furthermore it can be seen from energy considerations that all molecules which emerge from the source with radial velocity v_r less than $v_{\max}$, where $\frac{1}{2}mv_{\max}^2 = W(r=0.5\text{ cm}) - W(r=0)$ are held within the focuser by the electric field, while all molecules whose radial velocity is greater than $v_{\max}$ collide with the electrodes. Since $v_{\max}$ is a function of M_J (M_J is the projection of $\mathbf{J}$ on the direction of the electric field of the focuser) the number of molecules focused from a given zero-field quadrupole level depends on the high field distribution of these molecules among the various possible M states.

From kinetic theory, the number of molecules per second emerging from a thin-walled source of area S with radial velocity less than $v_{\max}$ is given by

$$N = PSv_0\Omega/(2\pi)^{\frac{1}{2}}kT, \quad (3)$$

where P is the source pressure, $v_0 = (kT/m)^{\frac{1}{2}}$ is the most probable velocity of molecules in the beam, T is the absolute temperature, and Ω is a solid angle defined by $\Omega = \pi(v_{\max}/v_0)^2$. The number of molecules per second in a given quadrupole level which are focused is therefore

$$N(F_1) = \frac{PSv_0}{(2\pi)^{\frac{1}{2}}kT} f(JKF_1) \sum_{M_J} \varphi(F_1 M_J)\Omega(M_J), \quad (4)$$

where $f(JKF_1)$ is the fraction of molecules emerging from the source in the quadrupole state characterized by J, K, and F_1 ($\mathbf{F}_1=\mathbf{J}+\mathbf{I}_N$, where I_N is the spin of the nitrogen nucleus), and $\varphi(F_1, M_J)$ is that fraction of these molecules which, according to the discussion in (I), go adiabatically into the state characterized by the quantum number M_J as they enter the high electric field of the focuser. The total number of molecules per second in the upper inversion state which are delivered to the cavity by the focuser is then just the sum of the $N(F_1)$ for the three quadrupole levels, and so is

$$N(J,K)=\frac{PSv_0}{(2\pi)^{\frac{1}{2}}kT}f(JK)\Omega(JK), \tag{5}$$

where

$$\Omega(JK)=\sum_{F_1}\frac{f(JKF_1)}{f(JK)}\sum_{M_J}\varphi(F_1M_J)\Omega(M_J)$$

is an average solid angle for the upper inversion state, and $f(JK)$ is the fraction of molecules emerging from the source in the upper inversion state of the JK rotational level.

If each of these N molecules could be induced to make a transition to the lower inversion state while in the resonant cavity the total power delivered by the beam would be just $N(JK)h\nu_0$. Actually only about 50 to 75% average transition probability for the molecules in the beam can be obtained due to the variation of their velocities and spatial orientations. Assuming 50% transition probability, a source temperature of 300°K, and the geometry and voltage of the focuser given above, a calculation of the solid angle for the 3-3 line gives $\Omega(3\text{-}3)=4\times10^{-3}$ steradian, and available power of 1.5×10^{-9} watt per square millimeter of source area at 1 mm Hg source pressure.

It is estimated that the total number of molecules emerging from the source in the solid angle from which the upper inversion state molecules are selected is about 10^{15} per second. This estimate comes from knowledge of the number of molecules necessary to induce oscillation. This indicates that the present source is operating fairly inefficiently.

RESONANT CAVITY AND LINE WIDTH

The beam of molecules which enters the resonant cavity is almost completely composed of molecules in the upper inversion state. During their flight through the cavity the molecules are induced to make downward transitions by the rf electric field existing in the cavity. The transition probability for any particular molecule at low field strengths is given from first-order perturbation theory by

$$P_{ab}=\hbar^{-2}|\mu_{ab}|^2\left|\int_0^{L/v}\mathcal{E}(t)e^{-i2\pi\nu_0 t}dt\right|^2, \tag{6}$$

where μ_{ab} is the dipole matrix element for the transition, L is the length of the cavity, v is the velocity of the molecule, and $\mathcal{E}(t)$ is the rf electric field at the position of the molecule.

An average transition probability $\bar{P}_{ab}$ can be obtained for all molecules in the beam by averaging over the various velocities, trajectories, and values of $|\mu_{ab}|$ for the molecules in the several states which contribute to each spectral line. The power emitted from the beam is then just

$$P=Nh\nu\bar{P}_{ab}, \tag{7}$$

so $\bar{P}_{ab}$ as a function of the frequency of the applied field determines the line width of the molecular response.

Under the simplifying assumptions that the molecules all travel axially down the length of the cavity, that their velocity is uniform and equal to v_0, and that the cavity is a perfect cylinder with only one-half wavelength of rf field in the axial direction, we find that the emitted power has a maximum at the natural transition frequency ν_B, and a total width at half-maximum of $1.2v_0/L$. If the field is assumed to be uniform along the axis rather than one-half of a sine wave, the corresponding total width at half-maximum is $0.9v_0/L$. This line width of about v_0/L can alternatively be obtained from the uncertainty principle and the finite time of interaction of the molecules with the rf field. Thus $\Delta\nu\approx1/\Delta t$, where Δt is the time of flight of the molecule in the cavity, or L/v_0. The identity of the "Doppler broadening" of the spectral line has essentially disappeared. The sharpness of the molecular response as opposed to that obtained in the usual spectrometer may alternatively be attributed to the long wavelength of the rf field in the cavity in the direction of travel of the beam. If the cavity is excited in a mode in which there is more than one-half wavelength in the direction of travel of the beam, then the molecular emission line, as given by Eq. (6), has two peaks symmetrically spaced about the transition frequency. The frequency separation of these peaks can be associated with the Doppler shift.

Equations (6) and (7) show that for small rf field strengths the emitted power P is proportional to $\mathcal{E}_{max}^2$ and thus to the energy stored in the cavity. For larger field strengths, of course, the molecular transitions begin to saturate and Eq. (6) is no longer sufficient for the calculation of P_{ab}. The effects of this saturation will be considered in detail in a later paper; however, we can say that P_{ab} must certainly be less than 1, and that if a high field strength is maintained in the cavity then the average transition probability $\bar{P}_{ab}(\nu_B)$ will be about 0.5. The total power available from the beam is therefore about $Nh\nu_B/2$. Power saturation in this case is rather similar to that in the usual molecular beam experiment, for which it has been considered by Torrey.[11]

Associated with the power emitted from the beam is an anomalous dispersion, that is, a sharp variation

[11] H. C. Torrey, Phys. Rev. **59**, 293 (1941).

in the dielectric constant of the cavity medium due to the beam. These two effects can be considered at the same time by thinking of the beam as a polarizable medium introduced into the cavity, whose average electric susceptibility is given by $\chi=\chi'+i\chi''$. The power emitted from the beam can then be shown directly from Maxwell's equations to be[12]

$$P=8\pi^2\nu_B W\chi'', \tag{8}$$

where W is the energy stored in the cavity. Thus, from Eqs. (7) and (8), χ'' is related to $\bar{P}_{ab}$ by

$$\chi''=Nh\bar{P}_{ab}/8\pi^2 W. \tag{9}$$

The value of χ' is given from χ'' by Kramer's relation,[13] which for a sharp resonance line can be approximated by[14]

$$\chi'(\nu)=\frac{1}{\pi}\int_0^\infty \frac{\chi''(\nu')d\nu'}{\nu'-\nu}. \tag{10}$$

Figure 2 shows the form of χ' and χ'', calculated with the assumptions that all molecules are traveling parallel to the axis of the cavity with uniform velocity, that the cavity is excited in the TE_{011} mode, and that there is a small field strength in the cavity so that $P_{ab}\ll 1$ and Eq. (6) is valid. χ' and χ'' can also be found directly by calculation of the induced dipole moments of the molecules as they traverse the cavity.

If the Q of the cavity is defined in terms of net power loss (i.e., $dW/dt=-2\pi\nu W/Q_C$) then the presence of the

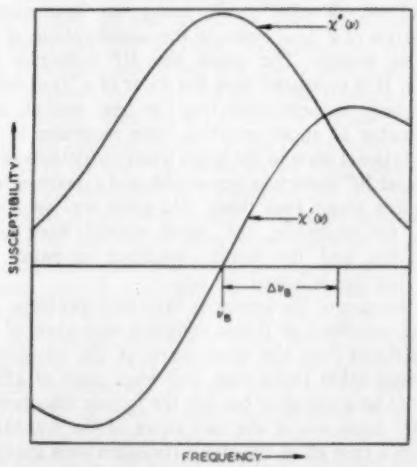


FIG. 2. Real and imaginary parts of the electric susceptibility χ of the molecular beam in the cavity. $\chi=\chi'+i\chi''$.

[12] J. C. Slater, *Microwave Electronics* (D. Van Nostrand Company, Inc., New York, 1950).
[13] J. H. Van Vleck, *Massachusetts Institute of Technology Radiation Laboratory Report*, 735, see also *Radiation Laboratory Series* (McGraw-Hill Book Company, Inc., New York, 1948), Vol. 13, Chap. 8.
[14] G. E. Pake and E. M. Purcell, Phys. Rev. **74**, 1184 (1948).

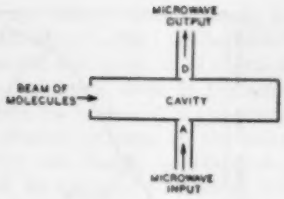


FIG. 3. Schematic diagram of the resonant cavity and molecular beam.

beam can be considered as causing a change in the effective Q_C given by $1/Q_{CB}=1/Q_C-4\pi\chi''$, where Q_{CB} and Q_C are respectively the cavity Q's with and without the beam, along with a shift in the resonant frequency of the cavity given by $\nu_{CB}=\nu_C(1-2\pi\chi')$ if $\chi'\ll 1$. These relations can also be easily derived directly from Maxwell's equations, and they will prove important in determining the properties of the maser.

THE MASER SPECTROMETER

Observed Line Shape as a Function of the Cavity Resonant Frequency

Consider the situation shown in Fig. 3. Power P_0 is incident on the cavity from wave guide A, and the power transmitted on out through wave guide D is detected as a function of the frequency of the input power. The power transmitted through the cavity in the absence of the beam is given by[12]

$$P_D(\nu)=\frac{P_0}{Q_A Q_D}\Big/\left[\left(\frac{1}{2Q_L}\right)^2+\left(\frac{\nu-\nu_C}{\nu_C}\right)^2\right], \tag{11}$$

where Q_A and Q_D are defined in terms of the power losses from the cavity to wave guides A and D respectively, and Q_L is the loaded Q of the cavity, given by $1/Q_L=1/Q_C+1/Q_A+1/Q_D$. As was shown in the last section, the change in $P_D(\nu)$ caused by the presence of the beam can be described through variations in Q_C and ν_C near the transition frequency ν_B. Thus in the presence of the beam we find P_D modified to

$$P_{DB}(\nu)=\frac{P_0}{Q_A Q_D}\Big/\left[\left(\frac{1}{2Q_L}-2\pi\chi''(\nu)\right)^2+\left(\frac{\nu-\nu_C+2\pi\nu_C\chi'(\nu)}{\nu_C}\right)^2\right]. \tag{12}$$

As long as the power output P_{DB} is not so high that nonlinearities in the molecular response are important, (12) gives the output power as a function of frequency in the presence of the beam and represents the spectrum which may be observed. For most spectroscopic applications we are interested in the case for which $\chi''(\nu)\ll 1/4\pi Q_L$ for all ν. For this case, an appropriate expansion of Eq. (12) shows that if $\nu_C=\nu_B$, where ν_B is the center frequency of the molecular response, then the presence of the beam shows up as a pip of the shape

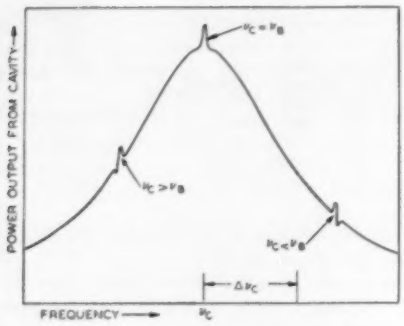


FIG. 4. Spectral line shapes as observed in the maser.

of χ'' superimposed on the cavity mode. The half-width $\Delta\nu_B$ of the molecular response is therefore defined as half the frequency separation of the half-maximum points of $\chi''(\nu)$. If the cavity frequency is altered so that the molecular line appears on the wings of the cavity mode, then the pip due to the beam assumes the shape of $\pm\chi'(\nu)$, the + or − sign depending on whether the line appears on the low or high frequency side of the cavity mode. These three situations are illustrated in Fig. 4.

Sensitivity of the Maser Spectrometer

We have here the usual problem occurring in microwave spectroscopy of detecting a small change in output power P_0 caused by the presence of the molecules. Assume that $\nu_C=\nu_B=\nu$ in Eq. (12). Then

$$P_{DB}=\frac{P_0}{Q_AQ_D}\Big/\left(\frac{1}{2Q_L}-2\pi\chi''(\nu_B)\right)^2. \tag{13}$$

The change in output power which must be detected is then

$$\delta P_D=P_{DB}-P_D\cong P_0(4Q_L^2/Q_AQ_D)\times[8\pi Q_L\chi''(\nu_B)]. \tag{14}$$

Consider now the change in output voltage δV_D, which is proportional to $(P_D+\delta P_D)^{\frac{1}{2}}-P_D^{\frac{1}{2}}$ or to $\delta P_D/2P_D^{\frac{1}{2}}$ when $\delta P_D\ll P_D$. The noise voltage at the output of a linear detector is just $(FkT\Delta\nu)^{\frac{1}{2}}$, where $\Delta\nu$ is the band width of the detector and F is its over-all noise figure. Thus the voltage signal-to-noise ratio is given by

$$\delta V_D/V_N=[(\delta P_D)^2/4FkTP_D\Delta\nu]^{\frac{1}{2}}. \tag{15}$$

As long as the change in power δP_D due to the beam increases linearly with the power level P_D, the signal to noise ratio from (15) continues to increase. Hence the power input to the cavity should be increased until the transition begins to saturate. For saturation and a given Q_C, it can be shown that maximum sensitivity is achieved by using a small input coupling ($Q_A\gg Q_C$) and a matched output coupling ($Q_D=Q_C$). If we approximate the saturation condition by setting W equal to the level at which the power emitted by the beam is just $\frac{1}{2}Nh\nu_B$, then with a little algebra we find, from Eqs. (8), (14), and the known relationship of W to P_0,[12] that the change in output power δP_D is just $\frac{1}{2}Nh\nu_B$ (the beam emits $\frac{1}{2}Nh\nu_B$ of power, and due to the change in the input match caused by the beam, $\frac{1}{2}Nh\nu_B$ more power enters the cavity through the input coupling hole A. Thus the increase in power input to the cavity is twice the power emitted from the beam. Half of this increase emerges into wave guide D since it was assumed to be matched to the cavity. The power level P_D is now determined by the required energy W from the relation

$$P_D=2\pi\nu_BW/Q_C. \tag{16}$$

This, with (8), gives

$$P_D=P/4\pi\chi''(\nu_B)Q_C=Nh\nu_B/8\pi\chi''(\nu_B)Q_C. \tag{17}$$

Inserting these values for P_D and δP_D in (15) yields

$$\delta V_D/V_N=[\pi Q_CNh\nu_B\chi''(\nu_B)/2FkT\Delta\nu]^{\frac{1}{2}}. \tag{18}$$

This relation gives the sensitivity of the maser, once the value of $\chi''(\nu_B)$ for a given molecule is calculated. χ'' is, of course, related to the average transition probability $\bar{P}_{ab}$ by Eq. (9), so that Eq. (18) can easily be rewritten in terms of the transition probability.

For the ammonia 3–3 line, a calculation of the number of molecules in the 3–3 state necessary to make $\delta V_D/V_N=1$ was done, assuming $F=100$, $\Delta\nu=1$ cps, $Q_C=12\,000$, $T=300°$ and using an approximate calculation of $\chi''(\nu_B)$ based on the considerations of the previous section. The result was 10^9 molecules per second. It is estimated from the value of $\chi''(\nu_B)$ which is necessary to cause oscillation (see next section) that the number of upper inversion state molecules in the 3–3 rotational state in the beam when oscillations occur is at least 10^{13} molecules per second, and experimentally a number about four times this great was achieved. Thus, for ammonia, the maser should have good sensitivity, and the results described in paper (I) show that this is indeed the case.

In the case of the ammonia inversion spectrum, the focuser can effect an almost complete separation of the upper states from the lower states of the transitions. For some other transitions, this ideal state of affairs may not be attainable, but yet the focuser may preferentially focus one of the two states of the transition. In such a case all of the above considerations apply so long as one uses for N just the excess number of molecules in one of the two states. It is, of course, unimportant for spectroscopic purposes whether the more highly focused state is the upper or lower state of the transition. The high sensitivity attained in the observation of the ammonia spectrum with the maser gives promise that it may be generally useful as a microwave spectrometer of very high resolution.

THE MASER OSCILLATOR AND AMPLIFIER

By extending the considerations of the previous section to include amplification of the thermal noise which exists in the cavity, we can discuss the properties of the maser as an oscillator or amplifier. The results of this analysis, which is made below, are as follows:

(1) The center frequency ν_0 of the oscillation is given to a good approximation by the equation

$$\nu_0=\nu_B+\frac{\Delta\nu_B}{\Delta\nu_C}(\nu_C-\nu_B), \tag{19}$$

where $\Delta\nu_C$ and $\Delta\nu_B$ are respectively the half-widths of the cavity mode and of the molecular emission line; and $\nu_C-\nu_B$ is the difference between the cavity resonant frequency ν_C and the line frequency ν_B.

(2) The total width at half-power of the spectral distribution of the oscillation is approximately

$$2\delta\nu=8\pi kT(\Delta\nu_B)^2/P_B \tag{20}$$

where T is the temperature and P_B is the power emitted from the beam. Inserting in (20) values which approximate the experimental conditions, $T=300°\text{K}$, $\Delta\nu_B=3\times10^3$ cps, $P_B=10^{-10}$ watt, we find $2\delta\nu\approx10^{-2}$ cps, or $\nu_B/\delta\nu=5\times10^{12}$.

(3) If the beam is sufficiently strong, the maser may be used as an amplifier with a gain greater than unity and a noise figure very close to unity.

The argument goes along the following lines. Consider the situation of Fig. 3, the cavity with two wave guides. The whole system will be assumed to be in thermal equilibrium in the absence of the beam. Noise power of amount kT per unit band width is incident on the cavity from each wave guide, and the cavity walls emit noise power within the cavity. Of the noise power incident on the cavity from wave guide A, a certain amount within the frequency range of the cavity mode enters the cavity; part of this power is then absorbed by the cavity walls and part is transmitted on out through wave guide D. A similar situation holds for noise power incident on the cavity from wave guide D; some is absorbed in the cavity and some is transmitted through to wave guide A. The cavity walls emit noise power in the region of the cavity mode and some of this power goes out through each wave guide. When the beam is not present we have assumed the system to be in thermal equilibrium, so there must be kT per unit band width of noise power flowing away from the cavity down each wave guide and there must be kT of noise energy in the cavity mode, as required by the equipartition theorem.

In the presence of the beam thermal equilibrium is upset. The beam, since it is composed solely of upper inversion state molecules, and since the probability for spontaneous decay of these molecules to the lower states is negligible during the time they take to traverse the cavity, contributes no random noise of its own to the rf field of the cavity. What it does is merely to amplify, in a way described by its effect on the loaded Q and resonant frequency of the cavity, all of the noise signals which exist in the cavity. Thus to the noise sources in the wave guides the intrinsic Q of the cavity seems to have been altered; whereas to the noise source within the cavity, the loading on the cavity seems to have changed. In fact, the presence of the beam can be duplicated in the imagination by attaching to the cavity a third wave guide, with a negative Q equal to $\frac{1}{4}\pi\chi''(\nu)$ describing its coupling to the cavity, and by simultaneously shifting the resonant frequency of the cavity by an amount $-2\pi\nu_C\chi'(\nu)$.

From these considerations we will show that in the presence of the beam more than kT of power per unit bandwidth travels down each wave guide away from the cavity. The extra power, of course, comes from the beam. At a certain critical beam intensity this power suddenly becomes large, corresponding to sustained oscillations.

Let $\Delta\nu$ be some arbitrarily small element of the frequency spectrum at frequency ν. Within this range noise power of magnitude $kT\Delta\nu$ is incident on the cavity from each wave guide, independent of ν. Let $P_A\Delta\nu$ be the amount of noise power which enters the cavity from the incident power in wave guide A, and let $P_A'\Delta\nu$ be the total noise power re-emitted into wave guide A from inside the cavity. The presence of the beam will be indicated by an added subscript; i.e., P_{AB} will represent the value of P_A when the beam is present, etc. Similar definitions apply to the output guide D. Since the noise powers generated in the wave guides and in the cavity are completely incoherent with one another, we can simply add power coming from various sources to obtain the total power in any element of the system. Thus the total energy $W\Delta\nu$ stored in the cavity is merely the sum $\sum_i W_i\Delta\nu$ of all the energies due to power coming from the various different power sources. (W is energy per unit band width.)

Consider now the flow of power when no beam is present. The noise power entering the cavity from wave guide A is given by[12]

$$P_A=\frac{kT}{Q_AQ_L^A}\Big/\left[\left(\frac{1}{2Q_L}\right)^2+\left(\frac{\nu-\nu_C}{\nu_C}\right)^2\right], \tag{21}$$

where $1/Q_L^s=\sum_{m\neq s}1/Q_m$ is just proportional to all the losses from the cavity except that due to Q_s. Of this power P_A, some is absorbed in the cavity walls, and the rest is transmitted on out through wave guide D. The energy per unit band width stored in the cavity due to this power input is

$$W_A=P_AQ_L^A/2\pi\nu$$
$$=\frac{kT}{2\pi\nu Q_A}\Big/\left[\left(\frac{1}{2Q_L}\right)^2+\left(\frac{\nu-\nu_C}{\nu_C}\right)^2\right], \tag{22}$$

and the power transmitted on to the output wave guide D is $2\pi\nu W_A/Q_D$, or $P_A Q_L{}^A/Q_D$. Similar expressions hold for the noise power incident on the cavity from wave guide D. Furthermore, the cavity loss associated with Q_C may be assumed to be due to a third wave guide with coupling characterized by Q_C to a perfectly conducting cavity, so that the energy W_C emitted into the actual cavity from its wall has the same form as (22). The total energy stored in the cavity per unit frequency interval is hence

$$W=W_A+W_D+W_C$$
$$=\frac{kT}{2\pi\nu Q_L}\Big/\left[\left(\frac{1}{2Q_L}\right)^2+\left(\frac{\nu-\nu_C}{\nu_C}\right)^2\right], \quad (23)$$

where

$$1/Q_L=1/Q_A+1/Q_D+1/Q_C.$$

The total energy stored in the cavity, given by $\int_0^\infty W d\nu$, is easily shown to be equal to kT (we make the assumption that $Q_L \gg 1$, so that in the integration the approximation $\nu \approx \nu_C$ may be made) as required by the equipartition theorem. The net noise power flowing in the wave guides A or D is also easily shown to be zero, so that the system is indeed seen to be in thermal equilibrium.

Consider now the case when the beam is present. The noise power incident from each wave guide sees a cavity whose rates of internal loss has been reduced by an amount $4\pi\chi''(\nu)$ by the energy emitted from the beam and whose resonant frequency has been shifted by an amount $-2\pi\nu_C\chi'(\nu)$. Corresponding to Eq. (21), the power entering the cavity from wave guide A in the presence of the beam is

$$P_{AB}=\frac{kT}{Q_{LB}{}^A Q_A}\Big/\left[\left(\frac{1}{2Q_{LB}}\right)^2+\left(\frac{\nu-\nu_{CB}}{\nu_{CB}}\right)^2\right], \quad (24)$$

where $1/Q_{XB}=1/Q_X-4\pi\chi''(\nu)$ for any x and $\nu_{CB}=\nu_C[1-2\pi\chi'(\nu)]$, while that entering from wave guide D is similarly

$$P_{DB}=\frac{kT}{Q_{LB}{}^D Q_D}\Big/\left[\left(\frac{1}{2Q_{LB}}\right)^2+\left(\frac{\nu-\nu_{CB}}{\nu_{CB}}\right)^2\right]. \quad (25)$$

The noise energy stored in the cavity due to these two sources is

$$W_{AB}=\frac{P_{AB}Q_{LB}{}^A}{2\pi\nu}, \quad \text{and} \quad W_{DB}=P_{DB}Q_{LB}{}^D/2\pi\nu. \quad (26)$$

At the same time, the energy stored in the cavity due to its own internal noise source is changed as though the loading on the cavity has been altered while its internal loss was unaffected. This energy is therefore given by

$$W_{CB}=\frac{kT}{2\pi\nu Q_C}\Big/\left[\left(\frac{1}{2Q_{LB}}\right)^2+\left(\frac{\nu-\nu_{CB}}{\nu_{CB}}\right)^2\right]. \quad (27)$$

Due to the presence of the beam the net noise power emitted from the cavity into the output wave guide D is now no longer zero. The power emerging from the cavity is now

$$P_{DB}'=\frac{2\pi\nu}{Q_D}[W_{CB}+W_{AB}]$$
$$=\frac{kT}{Q_D Q_L{}^D}\Big/\left[\left(\frac{1}{2Q_{LB}}\right)^2+\left(\frac{\nu-\nu_{CB}}{\nu_{CB}}\right)^2\right]. \quad (28)$$

Thus the additional noise output in wave guide D due to the beam is, from (25) and (28),

$$P_{DN}\equiv P_{DB}'-P_{DB}$$
$$=\frac{kT}{Q_D}4\pi\chi''(\nu)\Big/\left[\left(\frac{1}{2Q_{LB}}\right)^2+\left(\frac{\nu-\nu_{CB}}{\nu_{CB}}\right)^2\right]. \quad (29)$$

The power which must be emitted from the beam to give this amount of power in wave guide D is just

$$P_N=P_{DN}Q_D/Q_L=\frac{kT}{Q_L}4\pi\chi''(\nu)\Big/$$
$$\left[\left(\frac{1}{2Q_L}-2\pi\chi''(\nu)\right)^2+\left(\frac{\nu-\nu_C+2\pi\nu_C\chi'(\nu)}{\nu_C}\right)^2\right], \quad (30)$$

where ν_C has replaced ν_{CB} in the denominator of the last term in the denominator of this expression since $\nu_{CB}\approx\nu_C$. Note that (30) is just equivalent to (24) if the beam is thought of as a wave guide coupled to the cavity with a Q of $-1/4\pi\chi''$.

Expression (30) gives the complete spectrum of the power emitted from the beam due to amplification of the noise signals which are always present in the cavity. The necessary condition for the existence of oscillations as some cavity frequency is evidently that $\chi''(\nu_B)\approx\frac{1}{4}\pi Q_L$.

Assume that the cavity is tuned so that $|\nu_C-\nu_B| \ll \Delta\nu_C$, and then let the beam strength slowly increase so that χ'' increases. Then at the critical beam strength where $\chi''(\nu_B)\to 1/4\pi Q_L$, the total power $\int P_N d\nu$ emitted from the beam approaches infinity accordingly to (30). Obviously, the total power emitted from the beam cannot go to infinity, but is limited to about $\frac{1}{2}Nh\nu_B$. When the power level in the cavity reaches the point at which the molecular transition begins to saturate, χ'' and χ' become functions of the power level, and, of course, vary in such a way that $\int P_N d\nu$ is always less than $\frac{1}{2}Nh\nu_B$. We can, for simplicity, avoid the problem of dealing with this saturation merely by increasing χ'' until $\int P_N d\nu=\frac{1}{2}Nh\nu_B$ and examining the frequency spectrum of the power emitted from the beam at this level of output. Although χ'' is not independent of the electric field strength when saturation occurs, it varies much more slowly with time than does the oscillation, so that it may be

considered constant in treating the short-term behavior of the microwave field.

As the critical number of molecules is reached, P_N becomes very large at frequencies very close to ν_B. Hence it is appropriate to expand χ' and χ'' about the center frequency ν_B. This gives approximately

$$\chi'=\chi_0\left(\frac{\nu-\nu_B}{\Delta\nu_B}\right)+\cdots,$$
$$\chi''=\chi_0\left[1-\frac{1}{2}\left(\frac{\nu-\nu_B}{\Delta\nu_B}\right)^2+\cdots\right]. \tag{31}$$

Writing Eq. (30) in terms of (31), and setting $\int P_N d\nu$ equal to P_B, where $P_B=\frac{1}{2}Nh\nu_B$, one obtains

$$P_N\approx 4kT(\Delta\nu_B)^2\Big/\left[(\nu-\nu_0)^2+\left(\frac{4\pi kT}{P_B}(\Delta\nu_B)^2\right)^2\right], \tag{32}$$

where ν_0, the oscillation frequency, is given by the equation

$$\frac{\nu_0-\nu_B}{\nu_0-\nu_C}=-\frac{\Delta\nu_B}{\Delta\nu_C}, \tag{33}$$

or, as in (19),

$$\nu_0=\nu_B+(\nu_B-\nu_C)\Delta\nu_B/\Delta\nu_C \quad \text{if} \quad \Delta\nu_B/\Delta\nu_C\ll 1.$$

The total width $2\delta\nu$ at half-maximum power of this "noise" output is, from (32),

$$2\delta\nu=(8\pi kT/P_B)(\Delta\nu_B)^2 \tag{34}$$

as already stated in (20).

It should be remembered that (32) involves the assumption that the maser is a linear noise amplifier of very high gain. Actually, the noise properties of an oscillator depend to a considerable extent on the nonlinearities in its response, or the overload, and (32) does not accurately represent the precise noise spectrum of the maser as an oscillator. However, the approximate width of its noise spectrum is properly given by (34). As in the more usual types of oscillators, this oscillator actually maintains a nearly fixed amplitude of oscillation, but its phase slowly varies with time in a random way, corresponding to a noise spectrum of a width given approximately by (34). A more detailed discussion of noise will be given in a later publication.

The half-width of this oscillation signal is not to be confused with the half-width of the molecular response $\Delta\nu_B$. The latter represents the band width of the maser amplifier at low gain, whereas the former gives the band width of the oscillation signal. The oscillation frequency ν_0 can be varied throughout the range over which the molecules will amplify in accordance with (33) or (19). Hence care must be taken to keep the cavity frequency ν_C constant if it is desired to keep the oscillation frequency constant for any extended period of time.

Noise Figure and Band Width of the Amplifier

The noise figure of the maser amplifier may be easily found from the results of the foregoing sections. Assume that $\nu_C=\nu_B=\nu$, where ν is the frequency of the signal to be amplified. Also assume that the detector has a band width $\Delta\nu_{det}$ such that $\Delta\nu_{det}\ll\Delta\nu_B$. Equation (13) gives the signal power at the cavity output, while Eq. (29) gives the noise at the output in excess of kT. Thus we see that the signal-to-noise ratio at the output is just

$$\frac{P_0\Big/\left[Q_AQ_D\left(\frac{1}{2Q_L}-2\pi\chi_0\right)^2\right]}{kT\Delta\nu_{det}\left[1+\frac{4\pi\chi_0}{Q_D}\Big/\left(\frac{1}{2Q_L}-2\pi\chi_0\right)^2\right]}, \tag{35}$$

where $\chi''(\nu_B)=\chi_0$. At the input to the cavity, the signal to noise ratio is $P_0/kT\Delta\nu_{det}$. Therefore the noise figure F, which is just the ratio of these two quantities, is

$$F=Q_AQ_D\left[\frac{4\pi\chi_0}{Q_D}+\left(\frac{1}{2Q_L}-2\pi\chi_0\right)^2\right]. \tag{36}$$

At the same time, the power amplification available is, from (13), given simply by

$$\mu=P_{DB}/P_0=\left[Q_AQ_D\left(\frac{1}{2Q_L}-2\pi\chi_0\right)^2\right]^{-1}. \tag{37}$$

It can be shown from (37) that $\mu<1$ if $4\pi\chi_0<1/Q_C$, i.e., if there is a net loss of power within the cavity itself. Thus unless it is possible to produce oscillation by putting lossless reflections in all the wave guides so that $Q_L\to Q_C$, it is also impossible to create an amplifier with a gain greater than unity. In order to obtain a large gain μ, one must have $1/Q_L\approx 4\pi\chi_0$. If the gain is large, then a noise figure approaching unity is attainable by making $1/Q_A\approx+4\pi\chi_0\approx 1/Q_L$. This shows that for high amplification and at the same time a low noise figure, a fairly large input coupling to the cavity and a small output coupling is needed. Furthermore a sufficiently strong beam is required so that the maser is not too far from oscillation.

The maser acts as a regenerative amplifier, as can be seen from (12). Thus under conditions such that $4\pi\chi_0\approx 1/Q_L$ so that the midband gain is high, the band width becomes substantially smaller than $2\Delta\nu_B$.

It might also be noted that a certain amount of modulation of the amplified output is to be expected due to random variations of the number of molecules in the cavity at any time. These effects, however, are proportional to the input signal strength, and so are quite different from thermal noise signals which have no dependence on input power. Furthermore, they represent a modulation of only about one part in 10^6 since there are 10^{12} or more molecules in the cavity at

any time. This type of modulation can be neglected when small input signals are considered and is not important under most circumstances. This shot effect and also the effect of power flow through the cavity on the frequency dependence of the amplification will be discussed in more detail in a subsequent paper.

Amplification may also be accomplished using one wave guide as both input and output, and the noise figure of such an amplifier can also approach unity. The amplified output signal might be coupled out and detected through a directional coupler, which would have to have a fairly small coupling so that little of the input power was lost to it. Then so long as the amplified input noise appearing at the detector was large compared to kT, the noise figure of this amplifier would be small.

The maser amplifier may be useful in a restricted range of applications in spite of its narrow band width because of its potentially low noise figure. For example, suppose that the signal to be amplified came from outer space, where the temperature is only a few degrees absolute. Then by making the coupling through the cavity fairly large so that little noise is contributed by the cavity itself, amplification should be attainable while keeping the noise figure, *based on the temperature of the signal source*, fairly low. This might prove to have a considerable advantage over electronic amplifiers. It might also be possible to tune the frequency of a maser amplifier through the use of the Stark or Zeeman effects on the molecular transition frequencies.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the personnel of the Columbia Radiation Laboratory who assisted in the construction of the experimental apparatus. They would also like to thank Dr. T. C. Wang and Dr. Koichi Shimoda for their assistance and suggestions during the later stages of this work.

PHYSICAL REVIEW VOLUME 99, NUMBER 4 AUGUST 15, 1955

Resonance Transitions in Molecular Beam Experiments. I. General Theory of Transitions in a Rotating Magnetic Field*

HAROLD SALWEN
Columbia University, New York, New York
(Received March 29, 1955)

The time-dependent Schrödinger equation is solved formally for an atomic or molecular system which is subjected simultaneously to a rotating magnetic field of constant amplitude and angular velocity and to a constant magnetic field along the axis of rotation. The method yields the transition probabilities in terms of the solutions to an eigenvalue problem. This eigenvalue problem is solved both for (*a*) a normal Zeeman effect and for (*b*) the case where a transition from a given level is isolated in frequency from other transitions from the same level. Case (*a*) is exactly soluble and yields a solution which is shown to be the same as that of Bloch and Rabi, but is in a form which is more convenient for integration over the velocity distribution. Case (*b*) must be solved by an approximate method which results in a prediction of multiple quantum transitions as observed by Kusch.

IN the C-field of a molecular beams apparatus, a system (e.g., an atom or molecule) is subjected, in a region of constant magnetic field, to an oscillating magnetic field. The experimenter observes, as a function of frequency, a quantity proportional to the number of atoms which have left their initial states.

We will be dealing here with the case in which the constant and oscillating fields are perpendicular to each other. In this case, the solution is greatly simplified[1] by replacing the oscillating field by one which is rotating about the direction of the constant field. One must then correct for the effect of the other rotating component,[2] but the correction is usually small. In the present paper, we shall also explicitly assume that the constant magnetic field is uniform and that the rotating field is uniform in both amplitude and phase.

With these assumptions, one may, for a normal Zeeman effect, solve exactly for the probability of a transition being induced by the rotating field.[3] In other cases, however, some approximate method must be used. Allowed transitions may usually be treated by assuming that only the two levels involved interact,[4] but in the case of multiple quantum transitions which take place only by virtue of the existence of intermediate states, it is necessary to have a more general approach.

The approach we shall develop here has the virtue of not only yielding expressions for the probabilities of multiple quantum transitions but also giving corrections to the single quantum transition probabilities and a convenient expression for the exact solution of the normal Zeeman effect problem.

* Research supported by the National Science Foundation.
[1] See, for example, I. I. Rabi, Phys. Rev. **51**, 652 (1937).
[2] F. Bloch and A. Siegert, Phys. Rev. **57**, 522 (1940).
[3] F. Bloch and I. I. Rabi, Revs. Modern Phys. **17**, 237 (1945); E. Majorana, Nuovo cimento **9**, 43 (1932).
[4] H. C. Torrey, Phys. Rev. **59**, 293 (1941).

1. FORMAL SOLUTION

The problem may be stated as follows: A system whose unperturbed Hamiltonian is invariant under spatial rotations enters a region in which there is a constant field of magnitude H_z in the z-direction and a rotating field of amplitude H_0 in the xy-plane. If the initial state of the system is $|(t=0)\rangle$, what is the probability that at a time $t=\tau$ it will be observed to have left that state?

We now proceed to solve this problem.

A. System in a Constant Magnetic Field

Let $\mathcal{H}_0$ be the unperturbed Hamiltonian, $\mathbf{J}$ the total angular momentum, $\mathbf{J}_1$, $\mathbf{J}_2$, $\cdots\mathbf{J}_k$ the individual spin or orbital angular momenta which make up $\mathbf{J}$, and g_1, g_2, $\cdots g_k$ the corresponding gyromagnetic ratios. Since $\mathcal{H}_0$ is invariant under spatial rotations it will commute with both J^2 and J_z. Its eigenstates may then be taken to be eigenstates of J^2, J_z, and of a set of commuting variable, Γ, which commute with J^2 and J_z. The eigenvalues will be independent of J_z. The eigenstates will then satisfy the equation

$$\mathcal{H}_0|\gamma,j,m\rangle=W_0(\gamma,j)|\gamma,j,m\rangle, \tag{1}$$

where γ, $\hbar j(j+1)$, and $m\hbar$ are the eigenvalues of Γ, J^2, and J_z, respectively.

If we consider the effect of the constant field alone, the perturbed Hamiltonian will be

$$\mathcal{H}_1=\mathcal{H}_0+H_z(\mu_0/\hbar)[g_1J_{1z}+g_2J_{2z}+\cdots+g_kJ_{kz}]. \tag{2}$$

Although $\mathcal{H}_1$ in general no longer commutes with Γ or J^2 it does still commute with J_z. The steady states in this field will then be states for which m is a good quantum number but γ and j are not. For each state $|\gamma,j,m\rangle$ there will be one steady state which approaches it as H_z approaches zero. We will use the notation $|(\gamma,j),m\rangle$ for this state. The state $|(\gamma,j)m\rangle$ will then satisfy the equation

$$\mathcal{H}_1|(\gamma,j),m\rangle=W_1(\gamma,j,m)|(\gamma,j),m\rangle, \tag{3}$$

with

$$|(\gamma,j),m\rangle\rightarrow|\gamma,j,m\rangle \quad \text{as} \quad H_z\rightarrow 0, \tag{4}$$

where $W_1(\gamma,j,m)$ is the perturbed energy.

B. System with Rotating Field Included. Elimination of the Time from the Hamiltonian

Now we must introduce the rotating magnetic field. If the x-axis is chosen in the direction of the field at $t=0$ we have

$$H_x=H_0\cos\omega t, \quad H_y=H_0\sin\omega t. \tag{5}$$

Then the total Hamiltonian is

$$\begin{aligned}\mathcal{H}&=\mathcal{H}_1+H_x(\mu_0/\hbar)[g_1J_{1x}+g_2J_{2x}+\cdots+g_kJ_{kx}]\\&\quad+H_y(\mu_0/\hbar)[g_1J_{1y}+g_2J_{2y}+\cdots+g_kJ_{ky}]\\&=\mathcal{H}_1+\tfrac{1}{2}(H_x+iH_y)(\mu_1/\hbar)[g_1(J_{1x}-iJ_{1y})+\cdots\\&\quad+g_k(J_{kx}-iJ_{ky})]+\tfrac{1}{2}(H_x-iH_y)(\mu_0/\hbar)\\&\quad\times[g_1(J_{1x}+iJ_{1y})+\cdots+g_k(J_{kx}+iJ_{ky})],\\&=\mathcal{H}_1+H_0(\mu_0/2\hbar)\{e^{i\omega t}[g_1(J_{1x}-iJ_{1y})+\cdots\\&\quad+g_k(J_{kx}-iJ_{ky})]+e^{-i\omega t}[g_1(J_{1x}+iJ_{1y})+\cdots\\&\quad+g_k(J_{kx}+iJ_{ky})]\}.\end{aligned} \tag{6}$$

Making use of the fact[5] that $\langle\gamma,j,m|J_{kx}\pm iJ_{ky}|\gamma',j'm'\rangle$ is different from zero only when m' is equal to $m\mp1$, we find that

$$\begin{aligned}&e^{\mp i\omega t}\langle\gamma,j,m|J_{kx}\pm iJ_{ky}|\gamma',j',m'\rangle\\&=e^{-i m\omega t}\langle\gamma,j,m|J_{kx}\pm iJ_{ky}|\gamma',j',m'\rangle e^{im'\omega t}\\&=\langle\gamma,j,m|e^{-(i/\hbar)J_z\omega t}(J_{kx}\pm iJ_{ky})e^{(i/\hbar)J_z\omega t}|\gamma',j',m'\rangle,\end{aligned} \tag{7}$$

and, therefore

$$e^{\mp i\omega t}(J_{kx}\pm iJ_{ky})=e^{-(i/\hbar)J_z\omega t}(J_{kx}\pm iJ_{ky})e^{(i/\hbar)J_z\omega t}. \tag{8}$$

We thus obtain the expression

$$\begin{aligned}\mathcal{H}&=\mathcal{H}_1+H_0(\mu_0/2\hbar)e^{-(i/\hbar)J_z\omega t}\{[g_1(J_{1x}-iJ_{1y})+\cdots\\&\quad+g_k(J_{kx}-iJ_{ky})]+[g_1(J_{1x}+iJ_{1y})+\cdots\\&\quad+g_k(J_{kx}+iJ_{ky})]\}e^{(i/\hbar)J_z\omega t},\\&=e^{-(i/\hbar)J_z\omega t}\{\mathcal{H}_1+H_0(\mu_0/\hbar)\\&\quad\times[g_1J_{1x}+\cdots+g_kJ_{kx}]\}e^{(i/\hbar)J_z\omega t}\end{aligned} \tag{9}$$

since $\mathcal{H}_1$ commutes with J_z.

We may make use of the fact that the expression in the brackets does not contain the time to reduce the solution of our problem to the solution of a problem whose Hamiltonian is independent of the time.

If $|\ \rangle$ is a state of the system it will satisfy the time-dependent Schrödinger equation:

$$i\hbar(d/dt)|\ \rangle=\mathcal{H}|\ \rangle. \tag{10}$$

Then if we make the substitution

$$|'\rangle=e^{(i/\hbar)J_z\omega t}|\ \rangle, \tag{11}$$

$|'\rangle$ will satisfy the equation

$$\begin{aligned}&i\hbar(d/dt)|'\rangle\\&=e^{(i/\hbar)J_z\omega t}\{i\hbar(d/dt)|\ \rangle\}-\omega J_ze^{(i/\hbar)J_z\omega t}|\ \rangle\\&=e^{(i/\hbar)J_z\omega t}\mathcal{H}|\ \rangle-\omega J_ze^{(i/\hbar)J_z\omega t}|\ \rangle\\&=\{(\mathcal{H}_1-\omega J_z)+(\mu_0H_0/\hbar)\\&\quad\times[g_1J_{1x}+g_2J_{2x}+\cdots+g_kJ_{kx}]\}|'\rangle\\&=\mathcal{H}'|'\rangle,\end{aligned} \tag{12}$$

[5] See, for example, E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, London, 1951), Eq. 9'11.

where

$$\mathcal{H}'=(\mathcal{H}_1-\omega J_z)+(\mu_0 H_0/\hbar)[g_1 J_{1z}+\cdots+g_k J_{kz}] \quad (13)$$

is independent of the time.

C. Solution of the Equations of Motion. Calculation of Transition Probabilities[6]

Just as in any such problem, we may find the eigenstates of the operator by solving the secular equation:

$$\det(\mathcal{H}'-\lambda I)=0 \quad (14)$$

for the eigenvalues λ and then solving a set of linear equations for the corresponding state vectors $|\lambda\rangle$. In general, there will be one eigenvalue $\lambda_{\gamma j m}$ for each state $|\gamma,j,m\rangle$ or $|(\gamma,j),m\rangle$.

Now given any state $|\ \rangle$ at $t=0$, we may expand $|'\rangle$ in terms of the states $|\lambda\rangle$:

$$|'\rangle=\sum_\lambda|\lambda\rangle\langle\lambda|'\rangle, \quad (15)$$

where the coefficients $\langle\lambda|'\rangle$ will be functions of the time. Then the equation of motion

$$i\hbar(d/dt)|'\rangle=\mathcal{H}'|'\rangle, \quad (12)$$

becomes

$$\sum_\lambda|\lambda\rangle\{i\hbar(d/dt)\langle\lambda|'\rangle\}=\sum_\lambda\mathcal{H}'|\lambda\rangle\langle\lambda|'\rangle =\sum_\lambda|\lambda\rangle\lambda\langle\lambda|'\rangle, \quad (16)$$

or

$$i\hbar(d/dt)\langle\lambda|'\rangle=\lambda\langle\lambda|'\rangle, \quad (17)$$

with the solution

$$\langle\lambda|'\rangle=e^{-(i/\hbar)\lambda t}\langle\lambda|'(t=0)\rangle=e^{-(i/\hbar)\lambda t}\langle\lambda|(t=0)\rangle. \quad (18)$$

We finally obtain

$$|'\rangle=\sum_\lambda|\lambda\rangle e^{-(i/\hbar)\lambda t}\langle\lambda|(t=0)\rangle \quad (19)$$

or

$$|\ \rangle=e^{-(i/\hbar)J_z\omega t}\sum_\lambda|\lambda\rangle e^{-(i/\hbar)\lambda t}\langle\lambda|(t=0)\rangle. \quad (20)$$

We must now calculate the probability that a system initially in a state $|(t=0)\rangle$ has left that state by the time τ. We may write this as

$$P[|(t=0)\rangle\rightarrow]=1-|\langle(t=0)|(t=\tau)\rangle|^2, \quad (21)$$

where, by (20),

$$\langle(t=0)|(t=\tau)\rangle =\sum_\lambda\langle(t=0)|e^{-(i/\hbar)J_z\omega\tau}|\lambda\rangle e^{-(i/\hbar)\lambda\tau}\langle\lambda|(t=0)\rangle. \quad (22)$$

One ordinarily deals with initial states for which J_z is quantized. In this case, the initial state satisfies an equation of the form

$$J_z|(t=0)\rangle=\mu\hbar|(t=0)\rangle, \quad (23)$$

so that

$$\langle(t=0)|(t=\tau)\rangle=e^{-i\mu\omega\tau}\sum_\lambda|\langle(t=0)|\lambda\rangle|^2e^{-(i/\hbar)\lambda\tau}. \quad (24)$$

Then

$$\begin{aligned}P[&|(t=0)\rangle\rightarrow]\\ &=1-\textstyle\sum_{\lambda\lambda'}e^{(i/\hbar)(\lambda-\lambda')\tau}\\ &\qquad\times|\langle(t=0)|\lambda\rangle|^2|\langle(t=\mathrm{p})|\lambda'\rangle|^2\\ &=1-\textstyle\sum_{\lambda\lambda'}\cos[(1/\hbar)(\lambda-\lambda')\tau]\\ &\qquad\times|\langle(t=0)|\lambda\rangle|^2|\langle(t=0)|\lambda'\rangle|^2\\ &=1-\textstyle\sum_\lambda|\langle(t=0)|\lambda\rangle|^2\sum_{\lambda'}|\langle(t=0)|\lambda'\rangle|^2\\ &\quad+\textstyle\sum_{\lambda\lambda'}\{1-\cos[(1/\hbar)(\lambda-\lambda')\tau]\}\\ &\qquad\times|\langle(t=0)|\lambda\rangle|^2|\langle(t=0)|\lambda'\rangle|^2\\ &=4\textstyle\sum_{\lambda>\lambda'}|\langle(t=0)|\lambda\rangle|^2|\langle(t=0)|\lambda'\rangle|^2\\ &\qquad\times\sin^2[(1/2\hbar)(\lambda-\lambda')\tau]. \end{aligned} \quad (25)$$

D. Physical Interpretation of the Result

The above discussion may be interpreted in the following way.

The substitution (11) was a transformation from a fixed frame of reference to one which is rotating with angular velocity ω about the z-direction. In this frame of reference, the effective Hamiltonian is $\mathcal{H}'$ which is independent of the time. The steady states, $|\lambda\rangle$, of $\mathcal{H}'$ are thus the steady states of the system.

If the initial state is one of the eigenstates, the system will remain in this eigenstate and the transition probability will be zero. If, on the other hand, the eigenstate is a linear combination of two or more of the eigenstates, the final state will contain the same amount of each eigenstate, but the relative phases will be changed. This change in the phases will cause transitions to other linear combinations of the same eigenstates.

In the expression (25) for the transition probability, each term is the contribution due to the phase difference of a single pair of states, $|\lambda\rangle$ and $|\lambda'\rangle$. The effect of this phase difference is given by

$$4\sin^2[(1/2\hbar)(\lambda-\lambda')\tau]=|e^{-(i/\hbar)\lambda\tau}-e^{-(i/\hbar)\lambda'\tau}|^2. \quad (26)$$

The coefficient

$$\begin{aligned}&|\langle(t=0)|\lambda\rangle|^2|\langle(t=0)|\lambda'\rangle|^2\\ &\quad=(\text{initial probability of eigenstate }|\lambda\rangle)\\ &\qquad\times(\text{initial probability of eigenstate }|\lambda'\rangle)\end{aligned} \quad (27)$$

expresses the fact that the phase difference between these states can cause transitions only to the extent that *both* states are present initially.

[6] *Note added in proof.*—Mr. M. Hack has pointed out to me that Besset *et al.* [J. phys. radium **15**, 251 (1954)] similarly approach the problem of "multiple-flop" transitions in terms of the steady states of $\mathcal{H}'$. Mr. Hack has also independently applied this "spectral representation" to the problem of multiple quantum transitions (unpublished).

E. The Case of $|(t=0)\rangle = |(\gamma,j), m\rangle$. Selection Rule for Transitions

In the usual experiment, the system enters the radio-frequency field after it has already been in the constant magnetic field, so that our initial state should be taken to be one of the states $|(\gamma,j),m\rangle$. For this case, the formula for the transition probability becomes

$$P[|(\gamma,j),m\rangle\rightarrow]=4\sum_{\lambda<\lambda'}|\langle(\gamma,j),m|\lambda\rangle|^2 \times|\langle(\gamma,j),m|\lambda'\rangle|^2\sin^2[(1/2\hbar)(\lambda-\lambda')\tau]. \quad (28)$$

In Secs. 2 and 3, we shall derive the transition probabilities in the two cases which cover most experimental applications. It is nevertheless instructive to determine here the resonant frequencies which may generally be expected.

Unless an exact solution is possible, it is simplest to work in the representation in which $\mathcal{H}_1$ is diagonal. In this representation, we expand in terms of the $|(\gamma,j),m\rangle$, which also happen to be our initial states. According to the foregoing derivation, our effective Hamiltonian is

$$\mathcal{H}'=(\mathcal{H}_1-\omega J_z)+(\mu_0H_0/\hbar)[g_1J_{1z}+\cdots+g_kJ_{kz}], \quad (13)$$

with

$$\mathcal{H}_1=\mathcal{H}_0+(\mu_0H_z/\hbar)[g_1J_{1z}+\cdots+g_kJ_{kz}]. \quad (2)$$

The first term, $\mathcal{H}_1-\omega J_z$, of $\mathcal{H}'$ is completely diagonal in our chosen representation, while the remaining term is completely off-diagonal.

Under ordinary experimental conditions the amplitude of the oscillating field is much smaller than that of the constant field, so it is reasonable to treat the off-diagonal terms of $\mathcal{H}'$ (proportional to the rotating field amplitude H_0) as small perturbations. Thus, except when two or more levels are degenerate or nearly degenerate, the states $|(\gamma,j),m\rangle$ are good approximations to the eigenstates of $\mathcal{H}'$ and no transitions are to be expected.

We should now note that the diagonal elements

$$\langle(\gamma,j),m|\,\mathcal{H}'\,|(\gamma,j),m\rangle=W_1(\gamma,j,m)-m\hbar\omega \quad (29)$$

are linear functions of the angular frequency ω. The frequency may thus be chosen so as to produce degeneracy. Those frequencies at which degeneracy occurs will (for $H_0\ll H_z$) be the resonant frequencies. Our condition for degeneracy between the levels $|(\gamma,j),m\rangle$ and $|(\gamma',j'),m'\rangle$ is thus

$$\langle(\gamma,j),m|\,\mathcal{H}'\,|(\gamma,j),m\rangle=\langle(\gamma',j'),m'|\,\mathcal{H}'\,|(\gamma',j'),m'\rangle \quad (30)$$

or

$$(m-m')\hbar\omega=W_1(\gamma,j,m)-W_1(\gamma',j',m'). \quad (31)$$

At the rotational frequency given by Eq. (31), a resonance should occur involving transitions between the states $|(\gamma,j),m\rangle$ and $|(\gamma',j'),m'\rangle$.

Up till now we have spoken in terms of a rotating field, whose angular frequency can have either sign. For the actual oscillating field, the frequencies are given by[7]

$$|m-m'|\hbar\omega=|W_1(\gamma,j,m)-W_1(\gamma',j',m')|. \quad (32)$$

Resonances will be observed at the frequencies given by Eq. (32) only if the amplitude of the oscillating field is large enough to produce transitions in the available time.

2. APPLICATION TO NORMAL ZEEMAN EFFECT

A. Solution for Transition Probabilities

A normal Zeeman effect occurs when we may take $\mathcal{H}_1$ and $\mathcal{H}$ to be given by[8]

$$\mathcal{H}_1=\mathcal{H}_0+(g\mu_0H_z/\hbar)J_z,$$
$$\mathcal{H}=\mathcal{H}_1+(g\mu_0/\hbar)\{H_xJ_x+H_yJ_y\}. \quad (33)$$

In this case, we have

$$|(\gamma,j),m\rangle=|\gamma,j,m\rangle,$$
$$\mathcal{H}_1|\gamma,j,m\rangle=(E_0+mg\mu_0H_z)|\gamma,j,m\rangle, \quad (34)$$

and

$$\mathcal{H}'=\mathcal{H}_0+[(g\mu_0H_z/\hbar)-\omega]J_z+(g\mu_0H_0/\hbar)J_x. \quad (35)$$

Let us introduce the following notation:

$$a=g\mu_0H_z/\hbar,\quad b=g\mu_0H_0/\hbar,$$
$$\cos\varphi=(a-\omega)/[(a-\omega)^2+b^2]^{\frac{1}{2}},$$
$$\sin\varphi=b/[(a-\omega)^2+b^2]^{\frac{1}{2}}, \quad (36)$$
$$\theta=[(a-\omega)^2+b^2]^{\frac{1}{2}}\tau,$$

where τ is the time the system is in the oscillating field. Then

$$\mathcal{H}'=\mathcal{H}_0+(a-\omega)J_z+bJ_x$$
$$=\mathcal{H}_0+[(a-\omega)^2+b^2]^{\frac{1}{2}}[\cos\varphi J_z+\sin\varphi J_x] \quad (37)$$
$$=\mathcal{H}_0+[(a-\omega)^2+b^2]^{\frac{1}{2}}J_{z'},$$

where the z'-axis is as shown in Fig. 1. The eigenvalues of $J_{z'}$ are $\mu\hbar$ ($\mu=-j,\ -j+1,\ \cdots,\ j$). Then the eigenvalues and eigenvectors of $\mathcal{H}'$ will be

$$\lambda_{\gamma j\mu}=E_0(\gamma,j)+[(a-\omega)^2+b^2]^{\frac{1}{2}}\mu\hbar,$$
$$|\lambda_{\gamma j\mu}\rangle=|\gamma,j,J_{z'}=\mu\hbar\rangle, \quad (38)$$

[7] We should note there that, even though Eq. (31) includes all resonances due to a rotating field, Eq. (32) does not include all possible resonances, due to an oscillating field. An example of a resonance which is not predicted by (32) is that due to the two-quantum transition $a\leftrightarrow i\leftrightarrow b$ where $m_a=m_b$, $m_i=m_a\pm1$. In this case, the states $|a\rangle$ and $|b\rangle$ are connected by one rotating component, while the states $|a\rangle$ and $|b\rangle$ are connected by the opposite component. As a result there should be a resonance at the frequency $\hbar\omega=|(W_a-W_b)/2|$. This resonance will have a large peak intensity when $W_i\sim\frac{1}{2}(W_a+W_b)$.

[8] This may be true for several possible reasons: (1) All the g_k's may be equal. (2) Some of the J_k's may be identically equal to zero for the states under consideration and the remaining g_k's may be equal. (3) The magnetic field may be too weak to decouple certain of the angular momenta from each other with the result that Eq. (33) turns out to be a good approximation. Except in the first case, we should then take $g=g(\gamma,j)$.

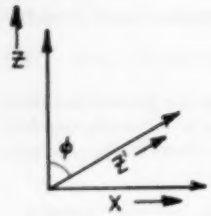


FIG. 1. Relation between the z- and z'-axes.

where the states $|\gamma,j,\mu\rangle$ are quantized along the z'-direction rather than the z-direction. The states $|\gamma,j,\mu\rangle$ may be obtained from the states $|\gamma,j,m\rangle$ by a rotation of the coordinate system through an angle φ. This rotation is effected by the irreducible representation of the rotation group corresponding to angular momentum j. In what follows, we shall use the notation of Wigner[9] for these operators.

The eigenfunctions of $\mathcal{H}'$ are given by

$$|\gamma,j,J_{z'}=\mu\hbar\rangle=\sum_m \mathfrak{D}^{(j)}(0,\varphi,0)_{\mu m}|\gamma,j,J_z=m\hbar\rangle. \quad (39)$$

We then have

$$\langle\gamma,\ j,\ J_z=m\hbar|\gamma,\ j,\ J_{z'}=\mu\hbar\rangle = \mathfrak{D}^{(j)}(0,\varphi,0)_{\mu m},$$
$$\langle\gamma,\ j,\ J_{z'}=\mu\hbar|\gamma,\ j,\ J_z=m\hbar\rangle = \mathfrak{D}^{(j)}(0,-\varphi,0)_{m\mu}. \quad (40)$$

The transition probabilities may now be found immediately:

$$P[|\gamma,j,m\rangle\to]=4\sum_{\mu>\mu'}|\mathfrak{D}^{(j)}(0,\varphi,0)_{\mu m}|^2 \times|\mathfrak{D}^{(j)}(0,\varphi,0)_{\mu' m}|^2\sin^2\tfrac{1}{2}(\mu-\mu')\theta. \quad (41)$$

B. Proof of Equivalence with the Result of Bloch and Rabi

In the paper of Bloch and Rabi,[10] the probability of a transition from $|\gamma,j,m\rangle$ to $|\gamma,j,m'\rangle$ is given by $|T_{mm'}|^2$, where (see Appendix of this paper)

$$T_{mm'}=T_{m'm}=\mathfrak{D}^{(j)}(R)_{mm'},$$
$$\mathfrak{D}^{(\frac{1}{2})}(R)=\begin{pmatrix}\cos(\theta/2)+i\cos\varphi\sin(\theta/2) & i\sin\varphi\sin(\theta/2)\\ i\sin\varphi\sin(\theta/2) & \cos(\theta/2)-i\cos\varphi\sin(\theta/2)\end{pmatrix}. \quad (42)$$

The same probability is given in the present theory by $|\langle\gamma,j,m'|(t=\tau)\rangle|^2$, where $|(t=0)\rangle=|\gamma,j,m\rangle$. Using Eq. (20), we get [see Eq. (113)]

$$\begin{aligned}\langle\gamma,j,m'|(t=\tau)\rangle&=\sum_\mu\langle\gamma,j,m'|\exp[-(i/\hbar)J_z\omega\tau]|\gamma,j,\mu\rangle\exp[-(i/\hbar)\{E_0+\mu\hbar[(a-\omega)^2+b^2]^{\frac{1}{2}}\}\tau]\langle\gamma,j,\mu|\gamma,j,m\rangle\\ &=\exp\{-(i/\hbar)[E_0+m'\hbar\omega]\tau\}\sum_\mu\mathfrak{D}^{(j)}(0,\varphi,0)_{\mu m'}e^{-i\mu\theta}\mathfrak{D}^{(j)}(0,-\varphi,0)_{m\mu}\\ &=\exp\{-(i/\hbar)[E_0+m'\hbar\omega]\tau\}\sum_\mu\mathfrak{D}^{(j)}(0,\varphi,0)_{\mu m'}\mathfrak{D}^{(j)}(0,-\varphi,-\theta)_{m\mu}\\ &=\exp\{-(i/\hbar)[E_0+m'\hbar\omega]\tau\}\mathfrak{D}^{(j)}(S)_{mm'},\end{aligned} \quad (43)$$

where

$$\begin{aligned}\mathfrak{D}^{(\frac{1}{2})}(S)&=\mathfrak{D}^{(\frac{1}{2})}(0,-\varphi,-\theta)\mathfrak{D}^{(\frac{1}{2})}(0,\varphi,0)\\ &=\begin{pmatrix}\cos(\varphi/2)e^{-i\theta/2} & -\sin(\varphi/2)e^{i\theta/2}\\ \sin(\varphi/2)e^{-i\theta/2} & \cos(\varphi/2)e^{i\theta/2}\end{pmatrix}\begin{pmatrix}\cos(\varphi/2) & \sin(\varphi/2)\\ -\sin(\varphi/2) & \cos(\varphi/2)\end{pmatrix}\\ &=\begin{pmatrix}\sin^2(\varphi/2)e^{i\theta/2}+\cos^2(\varphi/2)e^{-i\theta/2} & -\sin(\varphi/2)\cos(\varphi/2)(e^{i\theta/2}-e^{-i\theta/2})\\ -\sin(\varphi/2)\cos(\varphi/2)(e^{i\theta/2}-e^{-i\theta/2}) & \cos^2(\varphi/2)e^{i\theta/2}+\sin^2(\varphi/2)e^{-i\theta/2}\end{pmatrix}\\ &=\begin{pmatrix}\cos(\theta/2)-i\cos\varphi\sin(\theta/2) & -i\sin\varphi\sin(\theta/2)\\ -i\sin\varphi\sin(\theta/2) & \cos(\theta/2)+i\cos\varphi\sin(\theta/2)\end{pmatrix}\\ &=[\mathfrak{D}^{(\frac{1}{2})}(R)]^{-1}=\mathfrak{D}^{(\frac{1}{2})}(R^{-1}).\end{aligned} \quad (44)$$

Equation (42) was used to obtain the last line.

Then $\mathbf{S}=\mathbf{R}^{-1}$ and

$$|\langle\gamma,\ j,\ m|(t=\tau)\rangle|^2=|\mathfrak{D}^{(j)}(R^{-1})_{mm'}|^2=|T_{mm'}|^2, \quad (45)$$

which proves that our result is the same as that of Bloch and Rabi.

3. APPLICATION TO ANOMALOUS ZEEMAN SPLITTING AND TO TRANSITIONS FOR WHICH $(\gamma',j')\neq(\gamma,j)$

In the special case which we considered in Sec. 2, $W_1(\gamma,j,m)$ was given by

$$W_1(\gamma,j,m)=E_0+(g\mu_0H_z/\hbar)m\hbar = E_0+m\hbar a. \quad (46)$$

[9] E. Wigner, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atmospektren* (Edwards Brothers, Inc., Ann Arbor, 1944), Chap. 15. See especially Eq. (27), p. 180 which corresponds to Eq. (113) here.

[10] F. Bloch and I. I. Rabi (reference 3).

Then in the $|(\gamma,j),m\rangle$ representation the diagonal elements of $\mathcal{H}'$ are

$$\begin{aligned}\langle(\gamma,j),m|\mathcal{H}'|(\gamma,j),m\rangle&=W_1(\gamma,j,m)-m\hbar\omega\\&=E_0(\gamma,j)+m\hbar[a(\gamma,j)-\omega]\end{aligned}\tag{47}$$

[$g=g(\gamma,j)$; see reference 8]. At the frequency $\omega=a(\gamma,j)$, all diagonal elements corresponding to (γ,j) will be equal. This frequency is the resonant frequency in the case of a linear Zeeman effect.

We will deal in this section with the case which is, in a sense, the opposite of the above. We will make the assumption that if ω is chosen so that

$$\begin{aligned}&\langle(\gamma,j),m|\mathcal{H}'|(\gamma,j),m\rangle\\&\qquad\sim\langle(\gamma',j'),m'|\mathcal{H}'|(\gamma',j'),m'\rangle,\end{aligned}\tag{48}$$

then at that frequency the difference between these two diagonal elements and any others is large compared to the off-diagonal elements of $\mathcal{H}'$. In other words, we assume that in terms of the diagonal elements of $\mathcal{H}'$ a given level is degenerate with no more than one other level at a given frequency.

A. Simplified Treatment

Before entering into the actual calculations for this case, we will attempt to show by a somewhat simplified discussion of the problem at hand why the particular method of attack used in the following sections was chosen. Certain expressions will be written down here without justification. The justification will appear in the following sections.

Consider the simplest case to which the method can be applied, that in which the interactions of two levels $m=\pm\frac{1}{2}$, are considered. $\mathcal{H}'$ will then take the form

$$\mathcal{H}'=\begin{pmatrix}W_{\frac{1}{2}}-\frac{1}{2}\hbar\omega & b\hbar\alpha\\ b\hbar\bar{\alpha} & W_{-\frac{1}{2}}+\frac{1}{2}\hbar\omega\end{pmatrix},\tag{49}$$

where b is proportional to the rf amplitude and α, which in this case is most conveniently taken as $\frac{1}{2}$, will in further examples be taken proportional to the matrix element of $(g_1J_{1z}+g_2J_{2z}+\cdots+g_kJ_{kz})$.

Since the secular equation for $\mathcal{H}'$ is quadratic, we may solve it exactly, obtaining the eigenvalues

$$\begin{aligned}\lambda_\pm&=\tfrac{1}{2}[(W_{\frac{1}{2}}-\tfrac{1}{2}\hbar\omega)+(W_{-\frac{1}{2}}+\tfrac{1}{2}\hbar\omega)]\\&\quad\pm\tfrac{1}{2}[(W_{\frac{1}{2}}-W_{-\frac{1}{2}}-\hbar\omega)^2+4b^2\hbar^2|\alpha|^2]^{\frac{1}{2}}\\&=\tfrac{1}{2}(W_{\frac{1}{2}}+W_{-\frac{1}{2}})\pm\tfrac{1}{2}\hbar[(\omega_0-\omega)^2+4b^2|\alpha|^2]^{\frac{1}{2}},\end{aligned}\tag{50}$$

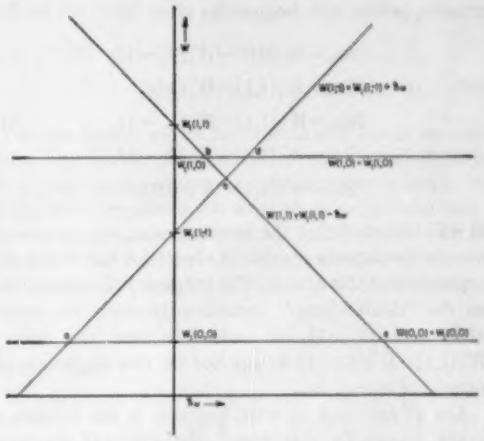


FIG. 2. Diagonal matrix elements as a function of frequency for a typical ordering of the energies $W_1(j,m)$. In this figure $W(j,m)$ is used as an abbreviation for

$$\langle(j),m|\mathcal{H}'|(j),m\rangle=W_1(j,m)-m\hbar\omega.$$

where the resonant frequency, ω_0, is given by $\hbar\omega_0=W_{\frac{1}{2}}-W_{-\frac{1}{2}}$. The corresponding eigenfunctions will have components

$$\begin{aligned}\langle\tfrac{1}{2}|\pm\rangle&=\frac{b\alpha}{|b\alpha|}\cdot\frac{1}{\sqrt{2}}\left[1\pm\frac{\omega_0-\omega}{[(\omega_0-\omega)^2+4b^2|\alpha|^2]^{\frac{1}{2}}}\right]^{\frac{1}{2}},\\\langle-\tfrac{1}{2}|\pm\rangle&=\pm\frac{1}{\sqrt{2}}\left[1\mp\frac{\omega_0-\omega}{[(\omega_0-\omega)^2+4b^2|\alpha|^2]^{\frac{1}{2}}}\right]^{\frac{1}{2}},\end{aligned}\tag{51}$$

so that, according to Eq. (25), the probability of a transition from one of the initial states $|\frac{1}{2}\rangle$ or $|-\frac{1}{2}\rangle$ is

$$\begin{aligned}P[|\tfrac{1}{2}\rangle\rightarrow]&=4|\langle\tfrac{1}{2}|+\rangle|^2|\langle\tfrac{1}{2}|-\rangle|^2\sin^2[(1/2\hbar)(\lambda_+-\lambda_-)\tau]\\&=\frac{4b^2|\alpha|^2}{(\omega_0-\omega)^2+4b^2|\alpha|^2}\\&\qquad\times\sin^2\{\tfrac{1}{2}[(\omega_0-\omega)^2+4b^2|\alpha|^2]^{\frac{1}{2}}\tau\}\\&=P[|-\tfrac{1}{2}\rangle\rightarrow].\end{aligned}\tag{52}$$

Now let us look at a more complicated case. If we allow only one value of γ but let j take the two values 1 and 0, $\mathcal{H}'$ will be of the form

$$\mathcal{H}'=\begin{bmatrix}W_1(1,1)-\hbar\omega & b\hbar\alpha(1,1;1,0) & 0 & b\hbar\alpha(1,1;0,0)\\ b\hbar\alpha(1,0;1,1) & W_1(1,0) & b\hbar\alpha(1,0;1,-1) & 0\\ 0 & b\hbar\alpha(1,-1;1,0) & W_1(1,-1)+\hbar\omega & b\hbar\alpha(1,-1;0,0)\\ b\hbar\alpha(0,0;1,1) & 0 & b\hbar\alpha(0,0;1,-1) & W_1(0,0)\end{bmatrix}\tag{53}$$

in the $|(\gamma,j),m\rangle$ representation. In Fig. 2, we have plotted the diagonal matrix elements of $\mathcal{H}'$ as a function of $\hbar\omega$, for a typical set of values of $W_1(1,1)$, $W_1(1,0)$, $W_1(1,-1)$, and $W_1(0,0)$. It is seen that there are five

crossing points with frequencies given by

$$\begin{aligned}
\hbar\omega_a &= W_1(0,0)-W_1(1,-1),\\
\hbar\omega_b &= W_1(1,1)-W_1(1,0),\\
2\hbar\omega_c &= W_1(1,1)-W_1(1,-1),\\
\hbar\omega_d &= W_1(1,0)-W_1(1,-1),\\
\hbar\omega_e &= W_1(1,1)-W_1(0,0).
\end{aligned} \tag{54}$$

It will be noted that the frequencies ω_a, ω_b, ω_d, and ω_e are the frequencies we should expect for the "allowed" transitions of this system. The frequency ω_c corresponds to the "double-jump" transition between the levels (1,1) and (1, −1), so called because the energy $W_1(1,1)-W_1(1,-1)$ is supplied by two equienergetic quanta of energy $\hbar\omega_c$.

Let us now look at what happens in the frequency range $\omega\sim\omega_b$. In this range, the diagonal elements $W_1(0,0)$ and $W_1(1,-1)+\hbar\omega$ are isolated from each other and from the other two diagonal elements, but $W_1(1,1)-\hbar\omega$ is about equal to $W_1(1,0)$. We may therefore expect $\mathcal{K}'$ to have two isolated roots

$$\begin{aligned}
\lambda_{0,0} &\sim W_1(0,0),\\
\lambda_{1,-1} &\sim W_1(1,-1)+\hbar\omega,
\end{aligned} \tag{55}$$

with

$$\begin{aligned}
|\lambda_{0,0}\rangle &\sim |(j=0),0\rangle,\\
|\lambda_{1,-1}\rangle &\sim |(j=1),-1\rangle,
\end{aligned} \tag{56}$$

and two almost degenerate roots

$$\lambda_{\pm} \sim W_1(1,1)-\hbar\omega \sim W_1(1,0), \tag{57}$$

with

$$|\lambda_{\pm}\rangle \sim A_{\pm}|(1),1\rangle + B_{\pm}|(1),0\rangle. \tag{58}$$

We would then expect to get a good approximation to $\lambda_{\pm}$ and $|\lambda_{\pm}\rangle$ by writing a two-by-two matrix in terms of $|(1),1\rangle$ and $|(1),0\rangle$, and including the effects of the other levels only as a perturbation on these. This argument suggests that a good approximation should be given by replacing the Hamiltonian by

$$\mathcal{K}' = \begin{pmatrix} W_1(1,1)+\Delta W(1,1)-\hbar\omega & b\hbar\alpha(1,1;1,0)+\Delta V \\ b\hbar\alpha(1,0;1,1)+\bar{\Delta}\bar{V} & W_1(1,0)+\Delta W(1,0) \end{pmatrix}, \tag{59}$$

with

$$\begin{aligned}
\Delta W(1,1) &= \frac{|b\hbar\alpha(1,1;0,0)|^2}{[W_1(1,1)-\hbar\omega]-W_1(0,0)} \sim \frac{|b\hbar\alpha(1,1;0,0)|^2}{\hbar(\omega_e-\omega_b)},\\
\Delta W(1,0) &= \frac{|b\hbar\alpha(1,0;1,-1)|^2}{W_1(1,0)-[W_1(1,-1)+\hbar\omega]} \sim \frac{|b\hbar\alpha(1,0;1,-1)|^2}{\hbar(\omega_d-\omega_b)},\\
\Delta V &= \frac{[b\hbar\alpha(1,1;0,0)][b\hbar\alpha(0,0;1,-1)][b\hbar\alpha(1,-1;1,0)]}{\{[W_1(1,1)-\hbar\omega]-W_1(0,0)\}\{[W_1(1,1)-\hbar\omega]-[W_1(1,-1)+\hbar\omega]\}}\\
&\sim \frac{[b\hbar\alpha(1,1;0,0)][b\hbar\alpha(0,0;1,-1)][b\hbar\alpha(1,-1;1,0)]}{\hbar(\omega_e-\omega_b)\cdot 2\hbar(\omega_c-\omega_b)}.
\end{aligned} \tag{60}$$

We may carry out the diagonalization of (59) as we did that of (49), obtaining for the dominant terms of $P[|(1),1\rangle\to]$ and $P[|(1),0\rangle\to]$

$$\begin{aligned}
P[|(1),1\rangle\to] &\sim P[|(1),0\rangle\to]\\
&\sim \frac{b^{*2}}{(\omega^*-\omega)^2+b^{*2}} \sin^2\{\tfrac{1}{2}[(\omega^*-\omega)^2+b^{*2}]^{\frac{1}{2}}\tau\},
\end{aligned} \tag{61}$$

where

$$\begin{aligned}
\hbar\omega^* &= \hbar\omega_b+\Delta W(1,1)-\Delta W(1,0),\\
b^* &= 2|b\alpha(1,1;1,0)+(1/\hbar)\Delta V|.
\end{aligned} \tag{62}$$

In the case shown in Fig. 2, we would expect this line to be shifted toward lower frequencies since $\Delta W(1,1)$ and $\Delta W(1,0)$ are both positive, with $\Delta W(1,1)$ having a larger energy denominator than $\Delta W(1,0)$.

We may consider the case of $\omega\sim\omega_c$ in the same way. In this case, the diagonal elements $W_1(1,1)-\hbar\omega$ and $W_1(1,-1)+\hbar\omega$ will be almost equal and there will be two eigenstates which are approximated by linear combinations of $|(1),1\rangle$ and $|(1),-1\rangle$. For this case we write, as in Eqs. (59) through (62) above,

$$\mathcal{K}' = \begin{bmatrix} W_1(1,1) \\ +\Delta W(1,1)-\hbar\omega & \Delta V \\ \bar{\Delta}\bar{V} & W_1(1,-1) \\ & +\Delta W(1,-1)+\hbar\omega \end{bmatrix}, \tag{63}$$

and

$$\begin{aligned}
P[|(1),1\rangle\to] &\sim P[|(1),-1\rangle\to]\\
&\sim \frac{b^{*2}}{(\omega^*-\omega)^2+b^{*2}} \sin^2\{(2/2)[(\omega^*-\omega)^2+b^{*2}]^{\frac{1}{2}}\tau\},
\end{aligned} \tag{64}$$

with

$$\Delta W(1,1)=\frac{|b\hbar\alpha(1,1;1,0)|^2}{\hbar(\omega_b-\omega_c)}+\frac{|b\hbar\alpha(1,1;0,0)|^2}{\hbar(\omega_a-\omega_c)},$$

$$\Delta W(1,-1)=\frac{|b\hbar\alpha(1,-1;1,0)|^2}{\hbar(\omega_c-\omega_d)}+\frac{|b\hbar\alpha(1,-1;0,0)|^2}{\hbar(\omega_c-\omega_a)}, \tag{65}$$

$$\Delta V=\frac{[b\hbar\alpha(1,1;1,0)][b\hbar\alpha(1,0;1,-1)]}{\hbar(\omega_b-\omega_c)}+\frac{[b\hbar\alpha(1,1;0,0)][b\hbar\alpha(0,0;1,-1)]}{\hbar(\omega_c-\omega_a)},$$

and

$$2\hbar\omega^*=2\hbar\omega_c+\Delta W(1,1)-\Delta W(1,-1),\qquad 2b^*=2|(1/\hbar)\Delta V|. \tag{66}$$

In order to justify the approach taken here, we shall derive in the next section a perturbation method which is based upon an expansion in terms of two nearly degenerate levels.

B. Perturbation Theory[11]

Consider a situation in which two unperturbed levels $|a\rangle$ and $|b\rangle$ are almost degenerate, but their unperturbed energies are quite different from those of any other level. The two perturbed states corresponding to $|a\rangle$ and $|b\rangle$ will then be approximated by a linear combination of $|a\rangle$ and $|b\rangle$. We therefore write

$$|\ \rangle=|a\rangle\langle a|\ \rangle+|b\rangle\langle b|\rangle+\sum_n|n\rangle\langle n|\ \rangle, \tag{67}$$

where the index n runs over states other than $|a\rangle$ and $|b\rangle$. (The coefficients $\langle n|\ \rangle$ will then be small.) By substituting (67) into the Schrödinger equation

$$\mathcal{H}|\ \rangle=W|\ \rangle, \tag{68}$$

we obtain

$$\langle a|\mathcal{H}|a\rangle\langle a|\ \rangle+\langle a|\mathcal{H}|b\rangle\langle b|\ \rangle+\sum_n\langle a|\mathcal{H}|n\rangle\langle n|\ \rangle=W\langle a|\ \rangle, \tag{69a}$$

$$\langle b|\mathcal{H}|a\rangle\langle a|\ \rangle+\langle b|\mathcal{H}|b\rangle\langle b|\ \rangle+\sum_n\langle b|\mathcal{H}|n\rangle\langle n|\ \rangle=W\langle b|\ \rangle, \tag{69b}$$

$$\langle n|\mathcal{H}|a\rangle\langle a|\ \rangle+\langle n|\mathcal{H}|b\rangle\langle b|\ \rangle+\sum_{n'}\langle n|\mathcal{H}|n'\rangle\langle n'|\ \rangle=W\langle n|\ \rangle. \tag{69c}$$

In principle, Eq. (69c) determines the coefficients $\langle n|\ \rangle$ in terms of $\langle a|\ \rangle$, $\langle b|\ \rangle$, and W, the solution taking the form

$$\langle n|\ \rangle=[W-\langle n|\mathcal{H}|n\rangle]^{-1}\times\{A_n(W)\langle a|\ \rangle+B_n(W)\langle b|\ \rangle\}. \tag{70}$$

If we are dealing with a finite number of levels we may solve exactly for the coefficients $A_n(W)$ and $B_n(W)$. If we are dealing with an infinite number of levels, or if the finite number is too large to make an exact solution convenient, we may expand A_n and B_n in terms of the off-diagonal part of $\mathcal{H}'$. In any case, we may substitute Eq. (70) into Eq. (69a, b), thereby obtaining

$$\begin{pmatrix}\langle a|\mathcal{H}|a\rangle+V_{aa}(W) & V_{ab}(W)\\ V_{ba}(W) & \langle b|\mathcal{H}|b\rangle+V_{bb}(W)\end{pmatrix}\begin{pmatrix}\langle a|\ \rangle\\ \langle b|\ \rangle\end{pmatrix}=W\begin{pmatrix}\langle a|\ \rangle\\ \langle b|\ \rangle\end{pmatrix}, \tag{71}$$

where

$$V_{aa}(W)=\mathcal{V}_{aa}(W)=\sum_n\frac{\langle a|\mathcal{H}|n\rangle A_n(W)}{W-\langle n|\mathcal{H}|n\rangle},$$

$$V_{ab}(W)=\mathcal{V}_{ba}(W)=\langle a|\mathcal{H}|b\rangle+\sum_n\frac{\langle a|\mathcal{H}|n\rangle B_n(W)}{W-\langle n|\mathcal{H}|n\rangle}, \tag{72}$$

$$V_{bb}(W)=\mathcal{V}_{bb}(W)=\sum_n\frac{\langle b|\mathcal{H}|n\rangle B_n(W)}{W-\langle n|\mathcal{H}|n\rangle}.$$

Equation (71) provides us with an iterative method for determining the perturbed energies. When we insert an approximate value of W into the left hand side of (71) the problem reduces to that of diagonalizing a two-by-two matrix. This yields us two eigenvalues, $W_\pm$. Now, if we put *one* of these into the left hand side of (71) we obtain *two* roots again, one of which corresponds to the eigenvalue for which we are looking. By selecting the proper root and continuing this process one may obtain as good an approximation as one desires.

The expressions for $W_\pm$, $\langle a|\pm\rangle$, and $\langle b|\pm\rangle$ (with the proper root selected) are

$$W_\pm=\tfrac{1}{2}\{[\langle a|\mathcal{H}|a\rangle+\langle b|\mathcal{H}|b\rangle]+[V_{aa}(W_\pm)+V_{bb}(W_\pm)]\}\pm\tfrac{1}{2}\{[\delta W(W_\pm)]^2+4|V_{ab}(W_\pm)|^2\}^{\frac{1}{2}}, \tag{73}$$

and

$$\langle a|\pm\rangle=\frac{V_{ab}(W_\pm)}{|V_{ab}(W_\pm)|}\cdot\frac{N_\pm}{\sqrt{2}}\times\left[1\pm\frac{\delta W(W_\pm)}{\{[\delta W(W_\pm)]^2+4|V_{ab}(W_\pm)|^2\}^{\frac{1}{2}}}\right]^{\frac{1}{2}},$$

$$\langle b|\pm\rangle=\pm\frac{N_\pm}{\sqrt{2}}\left[1\mp\frac{\delta W(W_\pm)}{\{[\delta W(W_\pm)]^2+4|V_{ab}(W_\pm)|^2\}^{\frac{1}{2}}}\right]^{\frac{1}{2}} \tag{74}$$

[11] For related perturbation methods see J. H. Van Vleck, Phys. Rev. **33**, 467 (1929); O. M. Jordahl, Phys. Rev. **45**, 87 (1934); M. H. L. Pryce, Proc. Phys. Soc. (London) **A63**, 25 (1950); O. M. Jordahl attributes his method to Kramers, Proc. Amsterdam Acad. Sci. **35**, 1272 (1932).

where

$$\delta W(W_\pm)=[\langle a|\,\mathcal{H}|a\rangle+V_{aa}(W_\pm)]]-[\langle b|\,\mathcal{H}|b\rangle+V_{bb}(W_\pm)], \quad (75)$$

and $N_\pm$ is a normalization constant, which is determined by the equation

$$|\langle a|\pm\rangle|^2+|\langle b|\pm\rangle|^2+\sum_n|\langle n|\pm\rangle|^2=1. \quad (76)$$

Once Eq. (73) has been used to obtain an accurate value of $W_\pm$, we may use Eqs. (74) and (70), to get the corresponding eigenstates, $|\pm\rangle$.

C. Transition Probabilities Near a Resonance

The results of the preceding section should enable us, in any particular case, to calculate the transition probability to the desired accuracy. We shall now carry this calculation through to a first approximation for the states $|a\rangle=|(\gamma_a,j_a),m_a\rangle$ and $|b\rangle=|(\gamma_b,j_b),m_b\rangle$.

Put $\mathcal{H}'$ in place of $\mathcal{H}$ in the equations of Sec. 3. B and let W_0 be defined by

$$W_0=\tfrac{1}{2}\{[\langle a|\,\mathcal{H}'|a\rangle+\langle b|\,\mathcal{H}'|b\rangle]+[V_{aa}(W_0)+V_{bb}(W_0)]\}. \quad (77)$$

If we use W_0 as a zeroth approximation to $W_\pm$, Eqs. (73) and (74) now give

$$W_\pm=W_0\pm\tfrac{1}{2}\{[\delta W(W_0)]^2+4|V_{ab}(W_0)|^2\}^{\frac{1}{2}}, \quad (78)$$

and

$$\langle a|\pm\rangle=\frac{V_{ab}(W_0)}{|V_{ab}(W_0)|}\cdot\frac{N_\pm}{\sqrt{2}}\times\left[1\pm\frac{\delta W(W_0)}{\{[\delta W(W_0)]^2+4|V_{ab}(W_0)|^2\}^{\frac{1}{2}}}\right]^{\frac{1}{2}}$$
$$\langle b|\pm\rangle=\pm\frac{N_\pm}{\sqrt{2}}\left[1\mp\frac{\delta W(W_0)}{\{[\delta W(W_0)]^2+4|V_{ab}(W_0)|^2\}^{\frac{1}{2}}}\right]^{\frac{1}{2}}. \quad (79)$$

Then the term in the transition probability which is large near the resonance $a\leftrightarrow b$ is given to this approximation by

$$4|\langle a|+\rangle|^2|\langle a|-\rangle|^2\sin^2(1/2\hbar)(W_+-W_-)\tau \sim|N_+|^2|N_-|^2\frac{4|V_{ab}(W_0)|^2}{[\delta W(W_0)]^2+4|V_{ab}(W_0)|^2}\times\sin^2(1/2\hbar)\{[\delta W(W_0)]^2+4|V_{ab}(W_0)|^2\}^{\frac{1}{2}}\tau. \quad (80)$$

Since

$$\langle a|\,\mathcal{H}'|a\rangle=W_1(\gamma_a,j_a,m_a)-m_a\hbar\omega\equiv W_1(a)-m_a\hbar\omega,$$
$$\langle b|\,\mathcal{H}'|b\rangle=W_1(\gamma_b,j_b,m_b)-m_b\hbar\omega\equiv W_1(b)-m_b\hbar\omega. \quad (81)$$

$\delta W(W_0)$ is given by

$$\delta W(W_0)=(m_a-m_b)\hbar(\omega_{ab}-\omega)+[V_{aa}(\omega_0)-V_{bb}(\omega_0)] \quad (82)$$

where ω_{ab} is defined by

$$(m_a-m_b)\hbar\omega_{ab}=W_1(a)-W_1(b). \quad (83)$$

We must bear in mind that W_0 is a function of ω. Nevertheless, it is clear from the form of Eq. (82) that there is a frequency $\omega^*\sim\omega_{ab}$ such that $\delta W(W_0(\omega^*))=0$. The frequency ω^* is then the corrected resonant frequency and, if we make the approximation that changes in V_{aa}, V_{bb}, and V_{ab} in the vicinity of the resonance may be ignored, the line shape will be given by

$$P[|a\rangle\rightarrow]\sim P[|b\rangle\rightarrow]\sim\frac{|N_+|^2|N_-|^2\cdot4|(1/\hbar)V_{ab}(W_0(\omega^*))|^2}{(m_a-m_b)^2(\omega^*-\omega)^2+4|(1/\hbar)V_{ab}(W_0(\omega^*))|^2}\times\sin^2\tfrac{1}{2}\{(m_a-m_b)^2(\omega^*-\omega)^2+4|(1/\hbar)V_{ab}(W_0(\omega^*))|^2\}^{\frac{1}{2}}\tau. \quad (84)$$

In analogy with Torrey's Eq. (4),[4] we may, omitting the factor $|N_+|^2|N_-|^2\sim1$, put this in the form

$$P[|a\rangle\rightarrow]\sim P[|b\rangle\rightarrow]\sim\frac{b_{ab}^2}{(\nu_{ab}^*-\nu)^2+b_{ab}^2}\times\sin^2\{\pi|m_a-m_b|[(\nu_{ab}^*-\nu)^2+b_{ab}^2]^{\frac{1}{2}}\tau\}, \quad (85)$$

with

$$\nu_{ab}^*=\omega^*/2\pi \quad (86)$$

and

$$b_{ab}=2|V_{ab}|/(|m_a-m_b|\hbar). \quad (87)$$

It may be shown that in the neighborhood of a resonance $|N_+|^2|N_-|^2=1+O(b^2)$ and terms other than (80) in the transition probability are of the order of b^2. It may also be shown that both these corrections are independent of frequency except for fourth-order terms. Thus Eq. (84) should give a good approximation in all cases in which the widths of "allowed" transitions are small compared to their separations.

D. Approximate Values of b_{ab} and ν_{ab}^*

In Sec. 3. B we merely remarked that Eq. (69c) could be solved for the coefficients $\langle n|\ \rangle$ but we did not carry out that solution. In this section we shall obtain the solution to the lowest order in b (the measure of the rf amplitude) and use that solution to get approximate values for the constants b_{ab} and ν_{ab}^* of Eq. (85).

Before carrying this out it is convenient to specify the states $|n\rangle$, $|n'\rangle$, etc., by

$$|n\rangle=|(\gamma,j),m\rangle,$$
$$|n'\rangle=|(\gamma',j'),m'\rangle, \quad (88)$$
$$\cdots\cdots\cdots\cdots$$
$$|n^{(k)}\rangle=|(\gamma^{(k)},j^{(k)}),m^{(k)}\rangle,$$

in analogy with the notation for $|a\rangle$ and $|b\rangle$. We will then define such expressions as ω_{an}, $\omega_{nn'}$, and $W_1(n)$, as we did similar expressions in Eqs. (31) and (33).

Now note that

$$\mathcal{H}'=(\mathcal{H}_1-\omega J_z)+(\mu_0 H_0/\hbar)\times[g_1 J_{1z}+g_2 J_{2z}+\cdots+g_k J_{kz}], \quad (13)$$

so that

$$\begin{aligned}\langle n|\,\mathcal{H}'|n\rangle&=W_1(\gamma,j,m)-m\hbar\omega\equiv W_1(n)-m\hbar\omega,\\ \langle n|\,\mathcal{H}'|n'\rangle&=\langle n|\,(\mu_0 H_0/\hbar)\times[g_1 J_{1z}+\cdots+g_k J_{kz}]|n'\rangle \quad \text{for } n'\neq n,\end{aligned} \quad (89)$$

where $\langle n|\,(\mu_0 H_0/\hbar)[g_1 J_{1z}+\cdots+g_k J_{kz}|n'\rangle$ is zero unless $m=m'\pm 1$. Then by suitable definitions of b and $\alpha(n;n')$ we have

$$\begin{aligned}\langle n|\,\mathcal{H}'|n\rangle&=W_1(n)-m\hbar\omega,\\ \langle n|\,\mathcal{H}'|n'\rangle&=b\hbar\alpha(n;n') \quad \text{for } n'\neq n,\end{aligned} \quad (90)$$

where

$$\alpha(n;n')=0, \quad \text{unless } m=m'\pm 1. \quad (91)$$

We shall adopt the convention that b is proportional to the rf amplitude and independent of the states involved, while $\alpha(n;n')$ depends only on the states.

Equation (69c) now becomes

$$[W-\langle n|\,\mathcal{H}'|n\rangle]\langle n|\;\rangle=b\hbar\alpha(n;a)\langle a|\;\rangle+b\hbar\alpha(n;b)\langle b|\;\rangle+\sum_{m'=m\pm 1} b\hbar\alpha(n;n')\langle n'|\;\rangle, \quad (92)$$

where the summation is over states other than $|a\rangle$ and $|b\rangle$. Then, due to the form of Eq. (70), we obtain

$$\begin{aligned}A_n(W)&=b\hbar\alpha(n;a)+\sum_{m'=m\pm 1}\frac{b\hbar\alpha(n;n')}{W-\langle n'|\,\mathcal{H}'|n'\rangle}A_{n'}(W),\\ B_n(W)&=b\hbar\alpha(n;b)+\sum_{m'=m\pm 1}\frac{b\hbar\alpha(n;n')}{W-\langle n'|\,\mathcal{H}'|n'\rangle}B_{n'}(W).\end{aligned} \quad (93)$$

We thus note that

$$A_n(W)=\begin{cases}O(b^0) & \text{if } m=m_a\\ O(b^k) & \text{if } |m-m_a|=k\neq 0,\end{cases} \quad (94)$$

with similar relations for $B_n(W)$.

We shall approximate $A_n(W)$ and $B_n(W)$ by their lowest order terms:

$$A_n(W)=\begin{cases}b\hbar\alpha(n;a) & \text{for } |m-m_a|=1\\ b\hbar\displaystyle\sum_{|m'-m_a|=1}\frac{[b\hbar\alpha(n;n')][b\hbar\alpha(n';a)]}{W-\langle n'|\,\mathcal{H}'|n'\rangle} & \text{for } m=m_a\\ b\hbar\displaystyle\sum_{m'=m_a\pm(k-1)}\frac{b\hbar\alpha(n;n')}{W-\langle n'|\,\mathcal{H}'|n'\rangle}A_{n'}(W) & \text{for } m=m_a\pm k,\ k>1\end{cases} \quad (95a)$$

$$B_n(W)=\begin{cases}b\hbar\alpha(n;b) & \text{for } |m-m_b|=1\\ b\hbar\displaystyle\sum_{|m'-m_b|=1}\frac{[b\hbar\alpha(n;n')][b\hbar\alpha(n';b)]}{W-\langle n'|\,\mathcal{H}'|n'\rangle} & \text{for } m=m_b\\ b\hbar\displaystyle\sum_{m'=m_b\pm(k-1)}\frac{b\hbar\alpha(n;n')}{W-\langle n'|\,\mathcal{H}'|n'\rangle}B_{n'}(W). & \text{for } m=m_b\pm k,\quad k>1\end{cases} \quad (95b)$$

The lowest order terms in $V_{aa}(W)$, $V_{bb}(W)$, and $V_{ab}(W)$ are then (see Eq. (72)):

$$V_{aa}(W)=b^2\hbar^2\sum_{|m'-m_a|=1}\frac{|\alpha(n';a)|^2}{W-\langle n'|\,\mathcal{H}'|n'\rangle}+O(b^4), \quad (96a)$$

$$V_{bb}(W)=b^2\hbar^2\sum_{|m'-m'|=1}\frac{|\alpha(n';b)|^2}{W-\langle n'|\,\mathcal{H}|n'\rangle}+O(b^4), \quad (96b)$$

$$V_{ab}(W)=\begin{cases}b\hbar\alpha(a;b)+O(b^2) & \text{for } |m_a-m_b|=1\\ b^k\hbar^k\displaystyle\sum_{\substack{m'=m_a\pm 1\\ m''=m_a\pm 2\\ \cdots\\ m^{(k-1)}=m_a\pm(k-1)}}\frac{\alpha(a;n')\alpha(n';n'')\cdots\alpha(n^{(k-1)};b)}{[W-\langle n'|\,\mathcal{H}'|n'\rangle][W-\langle n''|W'|n''\rangle]\cdots[W-\langle n^{(k-1)}|\,\mathcal{H}'|n^{(k-1)}\rangle]}+O(b^{k+2}). & \text{for } m_b=m_a\pm k,\quad k>1\end{cases} \quad (96c)$$

Then $\delta W(W)$ becomes

$$\begin{aligned}\delta W(W)&=\langle a|\,\mathcal{H}'|a\rangle-\langle b|\,\mathcal{H}'|b\rangle+V_{aa}(W)-V_{bb}(W)\\ &=[W_1(a)-m_a\hbar\omega]-[W_1(b)-m_b\hbar\omega]+V_{aa}(W)-V_{bb}(W)\\ &=(m_a-m_b)\hbar(\omega_{ab}-\omega)+V_{aa}(W)-V_{bb}(W).\end{aligned} \quad (97)$$

Since ω^* is defined as the root of $\delta W(W_0(\omega^*))=0$, we now obtain

$$\begin{aligned}\omega^*&=\omega_{ab}\mp(1/\hbar k)\{V_{aa}(W_0(\omega^*))-V_{bb}(W_0(\omega^*))\}\\ &=\omega_{ab}+O(b^2)\\ &=\omega_{ab}\mp(1/\hbar k)\{V_{aa}(W_0(\omega_{ab}))-V_{bb}(W_0(\omega_{ab}))\}+O(b^4),\end{aligned} \quad (98)$$

where

$$\pm k=m_b-m_a; \quad k>0. \quad (99)$$

This leads to

$$\begin{aligned}W_0(\omega^*)&=\tfrac{1}{2}[\langle a|\,\mathcal{H}'|a\rangle+\langle b|\,\mathcal{H}'|b\rangle]\\&\quad+\tfrac{1}{2}[V_{aa}(W_0(\omega^*))+V_{bb}(W_0(\omega^*))]\\&=\tfrac{1}{2}\{[W_1(a)-m_a\hbar\omega^*]\\&\quad+[W_1(b)-m_b\hbar\omega^*]\}+O(b^2)\\&=\tfrac{1}{2}\{[W_1(a)-m_a\hbar\omega_{ab}]\\&\quad+[W_1(b)-m_b\hbar\omega_{ab}]\}+O(b^2)\\&=W_1(a)-m_a\hbar\omega_{ab}+O(b^2)\\&=W_1(b)-m_b\hbar\omega_{ab}+O(b^2).\end{aligned}\tag{100}$$

Then we have

$$\begin{aligned}W_0(\omega^*)-\langle n'|\,\mathcal{H}'|n'\rangle&=[W_1(a)-m_a\hbar\omega_{ab}]-[W_1(n')-m'\hbar\omega^*]+O(b^2)\\&=(m_a-m')\hbar(\omega_{an'}-\omega_{ab})+O(b^2)\\&=(m_b-m')\hbar(\omega_{bn'}-\omega_{ab})+O(b^2).\end{aligned}\tag{101}$$

Finally, we obtain

$$\omega^*=\omega_{ab}\mp(b^2/k)\left\{\sum_{m'=m_a-1}\frac{|\alpha(n';a)|^2}{\omega_{an'}-\omega_{ab}}+\sum_{m'=m_a+1}\frac{|\alpha(n';a)|^2}{\omega_{ab}-\omega_{an'}}-\sum_{m'=m_b-1}\frac{|\alpha(n';b)|^2}{\omega_{bn'}-\omega_{ab}}-\sum_{m'=m_b+1}\frac{|\alpha(n';b)|^2}{\omega_{ab}-\omega_{bn'}}\right\}+O(b^4)\tag{102}$$

and

$$V_{ab}(W_0(\omega^*))=\begin{cases}b\hbar\alpha(a;b)+O(b^2) & \text{for } k=1\\ b^k\hbar\displaystyle\sum_{\substack{m'=m_a\pm1\\ m''=m_a\pm2\\ \cdots\\ m^{(k-1)}=m_a\pm(k-1)}}\frac{\alpha(a;n')\alpha(n';n'')\cdots\alpha(n^{(k-1)};b)}{(\pm)(\omega_{ab}-\omega_{bn'})(\pm)2(\omega_{ab}-\psi_{bn''})\cdots(\pm)(k-1)(\omega_{ab}-\omega_{bn^{(k-1)}})}+O(b^{k+2}). & \text{for } k>1\end{cases}\tag{103}$$

The above results immediately yield

$$\nu_{ab}^*=\nu_{ab}\mp(1/k)(b/2\pi)^2\left\{\sum_{m'=m_a-1}\frac{|\alpha(n';a)|^2}{\nu_{an'}-\nu_{ab}}+\sum_{m'=m_a+1}\frac{|\alpha(n';a)|^2}{\nu_{ab}-\nu_{an'}}-\sum_{m'=m_b-1}\frac{|\alpha(n';b)|^2}{\nu_{bn'}-\nu_{ab}}-\sum_{m'=m_b+1}\frac{|\alpha(n';b)|^2}{\nu_{ab}-\nu_{bn'}}\right\}+O(b^4),\tag{104}$$

and

$$b_{ab}=\begin{cases}2(b/2\pi)\alpha(a;b)+O(b^2) & \text{for } k=1\\ (2/k!)(b/2\pi)^k\left|\displaystyle\sum_{\substack{m'=m_a\pm1\\ m''=m_a\pm2\\ \cdots\\ m^{(k-1)}=m_a\pm(k-1)}}\frac{\alpha(a;n')\alpha(n';n'')\cdots\alpha(n^{(k-1)};b)}{(\nu_{ab}-\nu_{bn'})(\nu_{ab}-\nu_{bn''})\cdots(\nu_{ab}-\nu_{bn^{(k-1)}})}\right|+O(b^{k+2}), & \text{for } k>1\end{cases}\tag{105}$$

where

$$\nu_{ab}\equiv(\omega_{ab}/2\pi),\ \text{etc.}\tag{106}$$

We have thus obtained values of the constants b_{ab} and ν_{ab}^* which, when substituted into Eq. (84), will give us the approximate line shape. It should be noted that these results are in essential agreement with those of Sec. 3. A.

4. INTEGRATION OVER THE VELOCITY DISTRIBUTION

All the results obtained up to this point refer to an atom or molecule which is in the field for a given length of time, τ. In a molecular beams experiment this time is the transit time of a particle through the rf field and is therefore dependent on the particle velocity.

For a beam of particles which have escaped through a hole in an oven, the velocity distribution is given by

$$dn\propto v^3\exp(-mv^2/2kT)dv.\tag{107}$$

The result of averaging Eq. (25) over the velocity distribution is then

$$\begin{aligned}\bar{P}[|(t=0)\rangle\rightarrow]&=8(m/2kT)^2\int_0^\infty v^3\exp(-mv^2/2kT)\\&\quad\times\sum_{\lambda>\lambda'}|\langle(t=0)|\lambda\rangle|^2|\langle(t=0)|\lambda'\rangle|^2\\&\qquad\times\sin^2[(1/2\hbar)(\lambda-\lambda')L/v]dv\\&=8\sum_{\lambda>\lambda'}|\langle(t=0)|\lambda\rangle|^2|\langle(t=0)|\lambda'\rangle|^2\\&\quad\times\int_0^\infty x^3\exp(-x^2)\sin^2[(1/2\hbar)(\lambda-\lambda')\\&\qquad\times L(m/2kT)^{\frac{1}{2}}/x]dx,\end{aligned}\tag{108}$$

where L is the distance the particle travels in the rf

field. If we define the integral $K(\beta)$ by

$$K(\beta)=\int_0^\infty x^3 \exp(-x^2)\sin^2(\beta/2x)dx, \quad (109)$$

we have

$$\bar{P}[|(t=0)\rangle\rightarrow]=8\sum_{\lambda>\lambda'}|\langle(t=0)|\lambda\rangle|^2|\langle(t=0)|\lambda'\rangle|^2 \times K[(1/\hbar)(\lambda-\lambda')(m/2kT)^{\frac{1}{2}}L]. \quad (110)$$

We thus see that whenever our initial assumptions about homogeneity are valid the average over the velocity distribution is given by the function $K(\beta)$. We have used the methods suggested by Torrey[4] to calculate this integral for the full range of β, and the results will be published in a subsequent paper.

5. CONCLUSIONS

We have shown that when both the constant and the oscillating (rotating) magnetic fields are homogeneous, the solution may be reduced to the solution of an eigenvalue problem. This problem has been solved both for the case of a normal Zeeman effect and for the case in which transitions from a given state are isolated from each other.

For a normal Zeeman effect, the result is in agreement with previously known solutions of the problem. Our result [Eq. (41)] has the advantage that it is in a form in which the integral over the velocity distribution may immediately be obtained in terms of a single (soon to be tabulated) function.

In the second case, an approximate solution has been found for the line shape of a resonance due to a multiple-quantum transition. This result, given in Eq. (85), is in the same form as the "Rabi flopping formula" for allowed transition, except for a factor $|m_a-m_b|$ in the argument of the $\sin^2$. Another way of putting this is to say that the transit time, τ, in the Rabi formula is replaced by $|m_a-m_b|\tau$. We must therefore replace the criterion

$$\Delta\nu\tau_0\sim 1 \quad (111)$$

for the natural width of the line, by

$$|m_a-m_b|\Delta\nu\tau_0\sim 1, \quad (112)$$

where τ_0 is the transit time for the most probable velocity in the oven. The consequent narrowing of "double-flop" and "triple-flop" lines has been observed by Kusch.[12]

The detailed comparison of the theory with experiment has been deferred to a future paper.

The author wishes to thank Professor P. Kusch for suggesting the problem and for his constant cooperation in the course of the work. He is also indebted to Professor H. M. Foley and Professor N. M. Kroll for many helpful discussions.

APPENDIX. RELATIONSHIP OF THE WIGNER COEFFICIENTS TO THE $T_{mm'}$ OF BLOCH AND RABI

Wigner's formula for the coefficients $\mathfrak{D}^{(j)}(\alpha,\beta,\gamma)_{\mu'\mu}$ is (Eq. 27, reference 9)

$$\mathfrak{D}^{(j)}(\alpha,\beta,\gamma)_{\mu'\mu}=\sum_{\chi=\mathrm{Max}\left\{\begin{smallmatrix}0\\ \mu-\mu'\end{smallmatrix}\right\}}^{\mathrm{Min}\left\{\begin{smallmatrix}j+\mu\\ j-\mu'\end{smallmatrix}\right\}}(-1)^\chi\left\{\frac{[(j+\mu)!(j-\mu)!(j+\mu')!(j-\mu')!]^{\frac{1}{2}}}{(j+\mu-\chi)!(j-\mu'-\chi)!\chi!(\chi+\mu'-\mu)!}\right\} \times\{e^{i\mu'\alpha}\cos^{2j+\mu-\mu'-2\chi}(\beta/2)\sin^{2\chi+\mu'-\mu}(\beta/2)e^{i\mu\gamma}\} \quad (113)$$

so that $\mathfrak{D}^{(\frac{1}{2})}$ is given by

$$\mathfrak{D}^{(\frac{1}{2})}(\alpha,\beta,\gamma)=\begin{pmatrix} e^{i\alpha/2}\cos(\beta/2)e^{i\gamma/2} & e^{i\alpha/2}\sin(\beta/2)e^{-i\gamma/2}\\ -e^{-i\alpha/2}\sin(\beta/2)e^{i\gamma/2} & e^{-i\alpha/2}\cos(\beta/2)e^{-i\gamma/2}\end{pmatrix}. \quad (114)$$

Let us denote the components of this matrix by the letters A, B, C, and D. That is,

$$\mathfrak{D}^{(\frac{1}{2})}(R)=\begin{pmatrix} A & B\\ C & D\end{pmatrix}, \quad (115)$$

where the coefficients must satisfy the conditions

$$C=-B^*, \quad D=A^*, \quad |A|^2+|B|^2=1. \quad (116)$$

We may then see by direct comparison that

$$\mathfrak{D}^{(j)}(R)_{\mu'\mu}=\sum_{\chi=\mathrm{Max}\left\{\begin{smallmatrix}0\\ \mu-\mu'\end{smallmatrix}\right\}}^{\mathrm{Min}\left\{\begin{smallmatrix}j+\mu\\ j-\mu'\end{smallmatrix}\right\}}\left\{\frac{[(j+\mu)!(j-\mu)!(j+\mu')!(j-\mu')!]^{\frac{1}{2}}}{(j+\mu-\chi)!(\chi+\mu'-\mu)!(j-\mu'-\chi)!}\right\}\{A^{j+\mu-\chi}B^{\chi+\mu'-\mu}C^\chi D^{j-\mu'-\chi}\}. \quad (117)$$

[12] P. Kusch, Phys. Rev. **93**, 1022 (1954). Multiple quantum transitions have also been reported in a magnetic resonance experiment by Brossel, Cagnac, and Kastler [J. phys. radium **15**, 6 (1954)] and in electric quadrupole resonance by V. Hughes and L. Grabner [Phys. Rev. **79**, 829 (1950)].

Bloch and Rabi (Eq. 30, reference 3) give

$$T_{mm'}=\sum_{\rho=\mathrm{Max}\left\{\begin{smallmatrix}0\\-(m+m')\end{smallmatrix}\right\}}^{\mathrm{Min}\left\{\begin{smallmatrix}j-m\\j-m'\end{smallmatrix}\right\}}\left\{\frac{[(j+m)!(j-m)!(j+m')!(j-m')!]^{\frac{1}{2}}}{(m+m'+\rho)!(j-m'-\rho)!(j-m-\rho)!\rho!}\right\}\{A^{m+m'+\rho}B^{j-m'-\rho}C^{j-m-\rho}D^{\rho}\},\tag{118}$$

where A, B, C, and D are the corresponding coefficients for $j=\frac{1}{2}$. If we change the summation index to $\chi=j-m-\rho$ we get

$$T_{mm'}=\sum_{\chi=\mathrm{Max}\left\{\begin{smallmatrix}0\\m'-m\end{smallmatrix}\right\}}^{\mathrm{Min}\left\{\begin{smallmatrix}j+m'\\j-m\end{smallmatrix}\right\}}\left\{\frac{[(j+m)!(j-m)!(j+m')!(j-m')!]^{\frac{1}{2}}}{(j+m'-\chi)!(\chi+m-m')!\chi!(j-m-\chi)!}\right\}\{A^{j+m'-\chi}B^{\chi+m-m'}C^{\chi}D^{j-m-\chi}\}$$
$$=D^{(j)}(R)_{mm'},\tag{119}$$

where (Bloch and Rabi, reference 3, page 243)

$$\mathfrak{D}^{(\frac{1}{2})}(R)=\begin{pmatrix}A & B\\ C & D\end{pmatrix}=\begin{bmatrix}\cos\tfrac{1}{2}\lambda t-i\dfrac{\omega-\omega_0}{\lambda}\sin\tfrac{1}{2}\lambda t & i\dfrac{gH}{\lambda}\sin\tfrac{1}{2}\lambda t\\ i\dfrac{gH}{\lambda}\sin\tfrac{1}{2}\lambda t & \cos\tfrac{1}{2}\lambda t+i\dfrac{\omega-\omega_0}{\lambda}\sin\tfrac{1}{2}\lambda t\end{bmatrix}.\tag{120}$$

If we make the substitutions

$$gH=g_J\mu_0H_0/\hbar=b,\quad \omega_0=a,\quad \lambda\equiv[(\omega-\omega_0)^2+(gH)^2]^{\frac{1}{2}}=[(a-\omega)^2+b^2]^{\frac{1}{2}},\tag{121}$$

for the constants in terms of those in Sec. 2, we obtain

$$\mathfrak{D}^{(\frac{1}{2})}(R)=\begin{pmatrix}\cos(\theta/2)+i\cos\varphi\sin(\theta/2) & i\sin\varphi\sin(\theta/2)\\ i\sin\varphi\sin(\theta/2) & \cos(\theta/2)-i\cos\varphi\sin(\theta/2)\end{pmatrix}.\tag{122}$$

Since $B=C$, the $T_{mm'}$ are symmetric and

$$T_{mm'}=T_{m'm}=\mathfrak{D}^{(j)}(R)_{mm'}.\tag{123}$$

Classical Calculation of Differential Cross Section for Scattering from a Coulomb Potential with Exponential Screening*

EDGAR EVERHART, GERALD STONE, AND R. J. CARBONE
Physics Department, University of Connecticut, Storrs, Connecticut

(Received April 28, 1955)

The potential energy function considered is $V=(Z_1Z_2e^2/r)\exp(-r/a)$, which approximately represents the potential between two atoms in collision taking into account the screening of the atomic electrons. Here the first factor is the Coulomb potential and the exponential factor contains a screening length a. It is shown first that a classical orbital calculation should give valid results under certain conditions in problems where ions with energies of many thousands of electron volts scatter from atoms. Calculated values of the impact parameters and differential cross sections are presented for all angles of scattering. These quantities are tabulated for a wide range of parameters corresponding to various degrees of screening.

1. VALIDITY OF A CLASSICAL SOLUTION

THE interaction between two atoms during a collision is approximately represented by the potential energy function,

$$V=(Z_1Z_2e^2/r)\exp(-r/a), \tag{1}$$

over a wide range of energies. The first factor is the Coulomb potential energy function between two nuclei of charges Z_1e and Z_2e. The exponential factor takes into account the electron screening, whose extent is measured by the screening length a. Bohr[1] has discussed this potential in some detail and has suggested the expression,

$$a=a_0/[Z_1^{\frac{1}{2}}+Z_2^{\frac{1}{2}}]^{\frac{1}{2}}, \tag{2}$$

for the screening length. Here a_0 is 0.53×10^{-8} cm, the radius of the first orbit in hydrogen. Mott and Massey[2] have pointed out that a classical calculation of differential cross section for scattering is valid when: (a) the deBroglie wavelength λ of the incident particle is negligible compared with any significant dimension of the scattering center, and when (b) the collision is well defined within the limitations of the uncertainty principle.

Condition (a) requires that λ be much smaller than the screening length a, and also much smaller than the collision diameter b, defined by

$$b=Z_1Z_2e^2/(\tfrac{1}{2}mv^2). \tag{3}$$

In this expression m is the reduced mass of the system, and v is the relative velocity of the collision. The length b, which varies inversely with the energy of the collision would be the distance of closest approach in a "head on" collision were there no screening. Since this length is also the diameter of the cross section of backward scattering in the absence of screening, it is a good measure of the size of the scattering center when b/a is small. Bohr[1] has shown for this potential that the second condition, (b), leads to a lower limit on the scattering angle, namely,

$$\theta^*\approx\lambda/2\pi a. \tag{4}$$

For angles greater than this, the classical solution is valid.

By way of illustration, numbers will be put into the above equations for a particular case. For the collision of 50-kev neon ions with argon atoms, one calculates $a=160\times10^{-11}$ cm, $b=78\times10^{-11}$ cm, $\lambda=0.42\times10^{-11}$ cm, and $\theta^*=0.026$ radian, or 0.15°. Thus both conditions for the validity of a classical solution are satisfied at all angles greater than 0.15° in this case. In fact, for most collisions between atoms in the energy range from about a hundred electron volts to hundreds of thousands of electron volts, the classical calculation of differential cross section is valid except at very small angles. Massey and Smith[3] calculated differential cross sections for 72-ev protons on argon targets, and for 110-ev protons on helium targets using classical methods. They used, however, the self-consistent field instead of the screened Coulomb field considered in the calculations presented here.

There has been no general quantum mechanical solution worked out for the potential of Eq. (1). The Born approximation solution for this potential is well known,[1,4] but when the appropriate validity criteria are examined,[1,3,5] the solution for particles heavier than electrons is found to be valid only for angles less than θ^*, or $\lambda/2\pi a$. Since the classical solution is valid for angles greater than this limit, it is seen that the two methods are valid in mutually exclusive angular ranges.

* This work was sponsored by the Office of Ordnance Research through the Watertown Arsenal Laboratories and the Springfield Ordnance District.

[1] N. Bohr, Kgl. Danske Videnskab. Selskab, Mat-fys. Medd. **18**, 8 (1948). Paragraphs 1.4, 1.5, 1.6, and 2.1 are particularly pertinent to the discussion here.

[2] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, London, 1949), second edition, Chap. VII, Secs. 4 and 5.

[3] H. S. W. Massey and R. A. Smith, Proc. Roy. Soc. (London) **A142**, 142 (1933).

[4] See, for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949).

[5] Everhart, Stone, and Carbone, Technical Report No. 2 to Office of Ordnance Research, April 20, 1954 (unpublished).

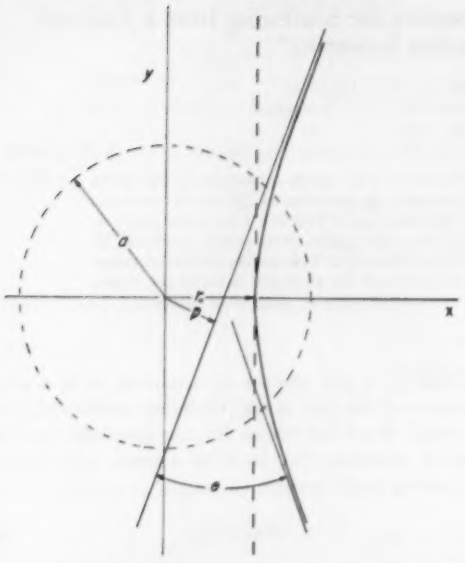


FIG. 1. Orbit of a particle scattered through an angle θ showing the impact parameter p, the distance of closest approach r_0, and the screening length a.

2. TABULATION OF RESULTS

Figure 1 shows a typical orbit in which the impact parameter p, the screening radius a, and the angle of scattering θ are indicated. The distance r_0 shown is the actual closest distance of approach for the orbit.

In any given experimental study, the particles in question and the energy will be known and the corresponding value of b/a is readily calculated from Eqs. (2) and (3). The case $b/a=0$ applies in the limit of higher energies where screening is negligible and here the potential reduces to the Coulomb potential. In this limit the differential cross section is given by the familiar Rutherford formula,

$$\sigma(\theta)=(b^2/16)\csc^4(\theta/2). \tag{5}$$

However, as the energy decreases, b/a increases, which corresponds to more effective screening.

The formulas for the classical calculation of the differential cross section at all angles from a given potential are well known. The integrations for the potential of Eq. (1) must be carried out numerically for each point. The formulas and details of the method of calculation are given in the Appendix to this paper. Of the two numerical procedures presented there, one method is suitable for all angles, and the other, somewhat simpler, is applicable at small angles only.

It is convenient to plot the ratio of the differential cross section $\sigma(\theta)$ to the square of the collision diameter b, since this dimensionless quantity is a function only of angle θ, and the ratio b/a. All calculations are in center-of-mass coordinates. Figure 2 shows $\sigma(\theta)/b^2$ plotted for angles up to 0.6π for eight values of b/a. It shows the general behavior of the screened differential cross section and its relationship to the Rutherford cross section.

Table I gives a complete summary of the numerical results, tabulating actual distance of closest approach, impact parameter, and differential cross section for angles between 0.01π and π and for $b/a=0$, 0.1, 0.2, 0.5, 1, 2, 5, 10. These calculations are particularly useful in interpreting results of experiments in which differential cross sections have been measured for single collisions between atoms at large angles.[6-8]

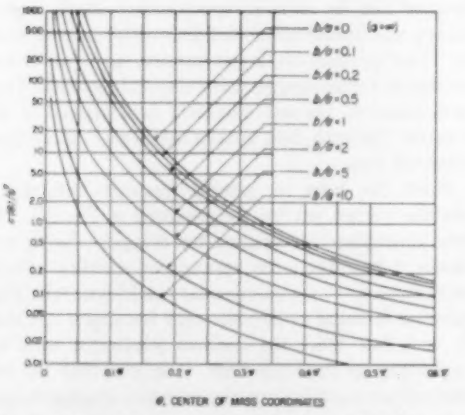


FIG. 2. Differential cross sections $\sigma(\theta)$ for scattering from an exponentially screened Coulomb potential plotted as a function of scattering angle θ in center-of-mass coordinates. Here b is the collision diameter and a is the screening length. For the case $b/a=0$, or $a=\infty$, there is no screening and the curve follows the Rutherford formula. Successively larger values of b/a correspond to more and more effective screening.

3. ACKNOWLEDGMENTS

We are pleased to acknowledge here our appreciation to Miss Alice Marconi and Mr. Hosy Lala who have carried out the painstaking numerical integrations.

4. APPENDIX

a. Large-Angle Calculations

The differential cross section for scattering from an arbitrary potential energy function $V(r)$ is calculated classically[2,9] from

$$\sigma(\theta)=-pdp/\sin\theta d\theta, \tag{6}$$

where

$$\theta=\pi-2\int_{r_0}^{\infty}[r\phi(r)]^{-1}dr, \tag{7}$$

[6] von C. Ramsauer and R. Kollath, Ann. Physik **16**, 570 (1933).
[7] A. G. Rouse, Phys. Rev. **52**, 1238 (1937).
[8] Everhart, Carbone, and Stone, Phys. Rev. **98**, 1045 (1955).
[9] H. S. W. Massey and E. H. S. Burhop, *Electronic and Ionic Impact Phenomena* (Oxford University Press, London, 1952), p. 373.

and

$$\phi(r)=[r^2/p^2-1-r^2V(r)/(\tfrac{1}{2}mv^2p^2)]^{\frac{1}{2}}. \quad (8)$$

Here r_0 is the largest positive root of Eq. (8). Letting $z=a/r$ in the above equations and substituting

$$V(r)=(\tfrac{1}{2}mv^2b/r)\exp(-r/a)$$

from Eqs. (1) and (3) into Eq. (7) there results

$$\theta=\pi-2(p/a)\int_0^{z_0} y^{-\frac{1}{2}}dz, \quad (9)$$

where

$$y=1-(p/a)^2z^2-(b/a)z\exp(-1/z), \quad (10)$$

and z_0 is the root of Eq. (10). The integral of Eq. (9) must be evaluated numerically for each value of p/a and b/a. In this form there is difficulty in maintaining sufficient accuracy at some angles because the second term of Eq. (9) is comparable with π and must then be subtracted from π to find θ. Letting

$$y_0=1-(p/a)^2z^2-(b/a)z\exp(-1/z_0), \quad (11)$$

Eq. (9) can be put into another form:

$$\theta=\pi-2(p/a)\left[\int_0^{z_0} y_0^{-\frac{1}{2}}dz-\int_0^{z_0}(y_0^{-\frac{1}{2}}-y^{-\frac{1}{2}})dz\right], \quad (12)$$

in which the first integral is readily evaluated analytically. The π cancels and the result is

$$\theta=2\cot^{-1}\left[\frac{2p/b}{\exp(-1/z_0)}\right]+\frac{2p}{a}\int_0^{z_0}(y_0^{-\frac{1}{2}}-y^{-\frac{1}{2}})dz. \quad (13)$$

In this form the numerical evaluation of the remaining integral by Simpson's rule yields a small quantity which is added to the first term and accuracy is maintained. There is an infinity in the integrand at $z=z_0$ which varies as $(z_0-z)^{-\frac{1}{2}}$ and has to be handled separately. Three place accuracy was maintained in the calculation of θ as a function of p/b.

The calculation of $dp/d\theta$ needed for Eq. (6) was done graphically using large scale plots of p/b *vs* θ on both linear and logarithmic scales. A careful plot of the resulting values of $dp/d\theta$ *vs* θ indicates that our values of this quantity for each point are accurate to within 2 percent. The final cross sections, calculated from Eq. (6) are thus correct to within 2 percent also. This should be sufficiently accurate for comparison with experimental data since the cross sections are rapidly varying functions of angle.

b. Small-Angle Calculations

At small angles there is a simpler method based on a calculation of the sideways impulse during a collision. As shown by the dotted line in Fig. 1, the actual path of the particle is approximated by a straight line $x=r_0$ parallel to the y-axis along which the particle moves with constant velocity v. Although the dotted line is

TABLE I. Distances of closest approach r_0, impact parameter p, and differential cross sections $\sigma(\theta)$ calculated classically for scattering from an exponentially screened Coulomb potential. Here a is the screening length and b is the collision diameter. Angles θ are in center-of-mass coordinates.

	$b/a=0$ $(a=\infty)$			$b/a=0.1$		
θ	r_0/b	p/b	$\sigma(\theta)/b^2$	r_0/b	p/b	$\sigma(\theta)/b^2$
0.01π	32.3	31.8	1.03×10^{4}	14.3	14.2	1.02×10^{3}
0.02π	16.4	15.9	6.42×10^{3}	9.85	9.65	1.47×10^{3}
0.03π	11.1	10.6	1.27×10^{3}	7.63	7.39	4.27×10^{2}
0.04π	8.46	7.95	4.02×10^{2}	6.28	6.00	1.71×10^{2}
0.06π	5.81	5.29	7.97×10^{1}	4.69	4.37	4.45×10^{1}
0.08π	4.49	3.96	2.53×10^{1}	3.78	3.42	1.62×10^{1}
0.10π	3.70	3.16	1.04×10^{1}	3.19	2.81	7.4×10^{0}
0.15π	2.64	2.08	2.11×10^{0}	2.35	1.91	1.67×10^{0}
0.20π	2.12	1.54	6.85×10^{-1}	1.90	1.43	5.7×10^{-1}
0.30π	1.60	0.981	1.47×10^{-1}	1.45	0.924	1.28×10^{-1}
0.40π	1.35	0.688	5.23×10^{-2}	1.24	0.651	4.64×10^{-2}
0.50π	1.21	0.500	2.50×10^{-2}	1.10	0.474	2.23×10^{-2}
0.75π	1.04	0.207	8.57×10^{-3}	0.961	0.197	7.7×10^{-3}
π	1.00	0	6.25×10^{-3}	0.913	0	5.65×10^{-3}
	$b/a=0.2$			$b/a=0.5$		
θ	r_0/b	p/b	$\sigma(\theta)/b^2$	r_0/b	p/b	$\sigma(\theta)/b^2$
0.01π	9.59	9.51	3.63×10^{3}	5.26	5.22	8.1×10^{2}
0.02π	7.11	6.99	6.3×10^{2}	4.16	4.10	1.70×10^{2}
0.03π	5.77	5.62	2.01×10^{2}	3.56	3.48	6.1×10^{1}
0.04π	4.94	4.75	8.8×10^{1}	3.14	3.04	2.91×10^{1}
0.06π	3.87	3.63	2.62×10^{1}	2.60	2.46	9.8×10^{0}
0.08π	3.19	2.92	1.06×10^{1}	2.22	2.06	4.32×10^{0}
0.10π	2.74	2.44	5.1×10^{0}	1.99	1.80	2.26×10^{0}
0.15π	2.09	1.73	1.29×10^{0}	1.59	1.35	6.5×10^{-1}
0.20π	1.73	1.32	4.65×10^{-1}	1.35	1.07	2.57×10^{-1}
0.30π	1.34	0.865	1.09×10^{-1}	1.08	0.729	6.9×10^{-2}
0.40π	1.13	0.615	4.06×10^{-2}	0.931	0.529	2.78×10^{-2}
0.50π	1.02	0.450	1.99×10^{-2}	0.839	0.391	1.43×10^{-2}
0.75π	0.878	0.187	7.0×10^{-3}	0.738	0.166	5.4×10^{-3}
π	0.845	0	5.14×10^{-3}	0.703	0	4.01×10^{-3}
	$b/a=1$			$b/a=2$		
θ	r_0/b	p/b	$\sigma(\theta)/b^2$	r_0/b	p/b	$\sigma(\theta)/b^2$
0.01π	3.22	3.20	2.80×10^{2}	1.91	1.90	8.5×10^{1}
0.02π	2.63	2.59	5.7×10^{1}	1.61	1.59	1.86×10^{1}
0.03π	2.30	2.25	2.15×10^{1}	1.43	1.40	7.2×10^{0}
0.04π	2.08	2.02	1.06×10^{1}	1.31	1.28	3.61×10^{0}
0.06π	1.78	1.70	3.82×10^{0}	1.15	1.10	1.34×10^{0}
0.08π	1.58	1.47	1.81×10^{0}	1.04	0.975	6.6×10^{-1}
0.10π	1.41	1.29	9.8×10^{-1}	0.957	0.880	3.76×10^{-1}
0.15π	1.16	1.00	3.13×10^{-1}	0.817	0.712	1.34×10^{-1}
0.20π	1.03	0.828	1.37×10^{-1}	0.723	0.595	6.3×10^{-2}
0.30π	0.844	0.588	4.25×10^{-2}	0.610	0.438	2.14×10^{-2}
0.40π	0.736	0.435	1.83×10^{-2}	0.542	0.332	1.00×10^{-2}
0.50π	0.671	0.326	9.8×10^{-3}	0.497	0.252	5.6×10^{-3}
0.75π	0.593	0.140	3.84×10^{-3}	0.440	0.110	2.36×10^{-3}
π	0.567	0	2.90×10^{-3}	0.426	0	1.83×10^{-3}
	$b/a=5$			$b/a=10$		
θ	r_0/b	p/b	$\sigma(\theta)/b^2$	r_0/b	p/b	$\sigma(\theta)/b^2$
0.01π	0.919	0.914	1.66×10^{2}	0.523	0.520	4.75×10^{1}
0.02π	0.798	0.789	3.63×10^{1}	0.460	0.455	1.06×10^{1}
0.03π	0.727	0.715	1.50×10^{1}	0.425	0.418	4.36×10^{0}
0.04π	0.679	0.662	7.8×10^{0}	0.398	0.389	2.31×10^{0}
0.06π	0.610	0.587	3.13×10^{0}	0.363	0.350	9.5×10^{-1}
0.08π	0.561	0.531	1.64×10^{0}	0.338	0.321	5.1×10^{-1}
0.10π	0.525	0.487	9.7×10^{-1}	0.319	0.299	3.15×10^{-1}
0.15π	0.465	0.411	3.80×10^{-1}	0.284	0.254	1.28×10^{-1}
0.20π	0.420	0.352	1.90×10^{-1}	0.263	0.222	6.8×10^{-2}
0.30π	0.360	0.267	7.1×10^{-2}	0.230	0.173	2.74×10^{-2}
0.40π	0.329	0.209	3.58×10^{-2}	0.211	0.138	1.48×10^{-2}
0.50π	0.305	0.162	2.17×10^{-2}	0.198	0.109	9.2×10^{-3}
0.75π	0.273	0.073	1.02×10^{-2}	0.180	0.050	4.74×10^{-3}
π	0.265	0	8.14×10^{-3}	0.175	0	3.84×10^{-3}

not close to the particular orbit shown in Fig. 1, it would lie close to the actual path of the particle if the figure were drawn for a case where the scattering angle θ was small. The x-component of the impulse along this path is assumed to be approximately the same as it would be along the actual path. This impulse over half the path is

$$mv_x=\int_0^\infty(-dV/dr)(x/r)dt, \tag{14}$$

where x/r is the cosine of the angle between the radial force $-dV/dr$ and the x-axis, and v_x is the final x-velocity. The integrand becomes a function of time alone when it is assumed that

$$x=r_0 \quad \text{and} \quad r=(r_0^2+v^2t^2)^{\frac{1}{2}}, \tag{15}$$

in accordance with the approximate path assumed above. An example of calculation of impulse in this way has been given by Bohr.[10] The ratio of the final x-momentum to the initial momentum mv is then half the angle of scattering. Thus

$$\theta/2=mv_x/mv. \tag{16}$$

For the potential energy function of Eq. (1), when one uses Eqs. (3), (14), (15), and (16), the integral is found to be

$$\frac{\theta a}{b}=\int_1^\infty\frac{(a/r_0+u)\exp(-ur_0/a)du}{u^2(u^2-1)^{\frac{1}{2}}}, \tag{17}$$

[10] See reference 1, p. 9.

where $u=(1+v^2t^2/r_0^2)^{\frac{1}{2}}$. Upon integrating the first term in the integrand of Eq. (17) by parts twice and the second term by parts once, the infinities at $u=1$ disappear and there is some cancellation. A simpler form results:

$$\theta a/b=(r_0/a)\int_1^\infty(u^2-1)^{\frac{1}{2}}\exp(-ur_0/a)du. \tag{18}$$

By letting $u=1/w$, an equivalent integral, with limits on w of zero to unity, can be obtained which is suitable for numerical integration. Although applicable only at small angles, this method has the advantage that $\theta a/b$ is a function only of r_0/a. Thus a single set of numerical integrations finds this function for all values of b/a.

Although r_0 is nearly equal to the impact parameter p at small angles, it is considerably more accurate not to assume them equal and to calculate p/r_0 for each value of b/a. By using Eq. (10) with $z_0=a/r_0$, the relationship is seen to be

$$p/r_0=[1-(b/r_0)\exp(-r_0/a)]^{\frac{1}{2}}. \tag{19}$$

By using Eqs. (18) and (19), it is possible to plot curves of p/b as a function of θ for each value of b/a. These curves agree with those obtained from the large-angle calculations of Eq. (13) very well at small angles, departing by no more than 1 percent at angles as large as 0.1π. The rest of the procedure, obtaining $dp/d\theta$ and calculating the differential cross section, is the same as has been described in the large-angle calculations.

Use of the Statistical Field Approximation in Molecular Physics

J. W. SHELDON*
Wilson Point, South Norwalk, Connecticut
(Received November 26, 1954)

The equations for the Dirac statistical field for an arbitrary molecule are obtained with boundary conditions, virial theorem, and an identity containing charge and energy integrals. Equations are obtained in which the forces between nuclei are related to kinetic energy. Numerical solutions of the statistical field equations for nitrogen are presented in graphical form, together with a discussion of numerical methods used to obtain them. The numerical results in conjunction with theoretical arguments show that it is not plausible that the statistical field should lead to binding for any molecule.

1. INTRODUCTION

IN 1932, Hund[1] obtained an approximate solution to the Thomas-Fermi equation for the ground state of the nitrogen molecule. He did not attempt to obtain the "Condon-Morse curve" or the dissociation energy of the molecule. In 1940, Eichelberger[2] obtained approximate solutions of the Thomas-Fermi equation for nitrogen for various internuclear distances. His solutions were solutions of difference equations approximating the Thomas-Fermi equation, but, as hand computing was used, were very rough because only a few mesh-points could be chosen. Eichelberger's calculations showed that the Thomas-Fermi equation, and Thomas-Fermi equation corrected for exchange, gave a stable nitrogen molecule. Although it was known that Eichelberger's results were very rough from a quantitative viewpoint, it was hoped that they were qualitatively correct, i.e., that the Thomas-Fermi equation led to a stable molecule. In August, 1952, solutions to difference equations approximating the Thomas-Fermi-Dirac equation for the nitrogen molecule were obtained on the IBM electronic data processing machines (Type 701 and associated equipment) located at the Poughkeepsie Laboratories of International Business Machines Company. The objectives of the calculations were to obtain points on the "Condon-Morse curve" for the molecule and to obtain a statistical field for nitrogen which could be used as a starting approximation for better models of the molecule and for other molecular calculations where a statistical field is satisfactory. Nitrogen was chosen because previous work has been done on this molecule. As a result of these calculations and from additional considerations here given, it has been possible to conclude, quite generally, that we may not expect a statistical field treatment for any molecule to give a stable molecule.

2. STATISTICAL FIELD FOR A GENERAL MOLECULE

A derivation of the statistical field equations with boundary conditions for a general molecule has not appeared previously. We give a brief derivation here, referring the reader to the literature for details of many steps (see, especially, Dirac[3] and Frankel[4]).

We consider the ground state of a system consisting of an even number E electrons in the field of M nuclei at rest. We assume that there are a number of orthonormal occupiable states $|a\rangle$, $|b\rangle$, $\cdots$, $|g\rangle$ equal in number to the number of electrons.[3] When electron i is in a particular state, say $|b\rangle$, we write $|b_i\rangle$. When electron i is in the ith of states $|a\rangle$, $|b\rangle$, $\cdots$, $|g\rangle$ we write $|\alpha_i\rangle$ or $|\beta_i\rangle$. Let

$$|Q\rangle=|a_1\rangle|b_2\rangle\cdots|g_E\rangle=|a_1b_2\cdots g_E\rangle. \tag{2.1}$$

$|Q\rangle$ is a possible state of the system and so is $P|Q\rangle$, where P is an operator representing a permutation of the electrons among the states. We apply the principle of superposition and the Pauli exclusion principle and take as the state of the system

$$|R\rangle=(1/\sqrt{n!})\textstyle\sum_P \epsilon_P P|Q\rangle, \tag{2.2}$$

where $\epsilon_P=+1$ for even permutations and -1 for odd permutations. The total energy of the system is

$$W=\langle R|H|R\rangle, \tag{2.3}$$

where

$$H=\textstyle\sum_i E_i+\sum_{i>j}F_{ij}+N. \tag{2.4}$$

E_i is the energy operator for electron i's kinetic energy plus its potential energy in the field of the nuclei, F_{ij} is the operator for the potential interaction energy between electrons i and j, and N is the total nuclear interaction energy. Because P commutes with H, because interactions take place between at most pairs of electrons, and because the one-electron states are orthonormal, we obtain (for details see Frankel[4])

$$W=\textstyle\sum_i\langle\alpha_i|E|\alpha_i\rangle+\frac{1}{2}\sum_{i,j}\langle\alpha_i\beta_j|F_{ij}\times\{|\alpha_i\beta_j\rangle-P_{ij}|\alpha_i\beta_j\rangle\}+N. \tag{2.5}$$

P_{ij} in (2.5) is a permutation of electrons i and j between states i and j. We have not more than two electrons i, j in any one term in W. We call these 1 and 2 and sum over states rather than electrons. Henceforth, α and β will indicate the "αth" and "βth" of states

* Presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Pure Science, Columbia University, New York, New York.

[1] F. Hund, Z. Physik 77, 12 (1932).

[2] J. F. Eichelberger, thesis, University of Ohio, 1940 (unpublished).

[3] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, London, 1947), third edition.

[4] J. Frankel, *Wave Mechanics, Advanced General Theory* (Oxford University Press, London, 1934), Chap. VIII.

$|a\rangle, |b\rangle, \cdots, |g\rangle$. Then

$$W=\sum_\alpha\langle\alpha_1|E_1|\alpha_1\rangle+\tfrac{1}{2}\sum_{\alpha,\beta}\langle\alpha_1\beta_2|F_{12}\times\{|\alpha_1\beta_2\rangle-|\alpha_2\beta_1\rangle\}+N. \quad (2.6)$$

Let

$$|\alpha\rangle=|\gamma s\rangle, \quad (2.7)$$

$$|\beta\rangle=|\delta t\rangle. \quad (2.8)$$

$|\gamma\rangle$, $|\delta\rangle$ are states belonging to an operator which is a function of the space coordinates only; $|s\rangle$, $|t\rangle$ states belonging to the spin energy operator. With (2.7), (2.8) in (2.6) we obtain, after summing over spins,

$$W=2\sum_\gamma\langle\gamma_1|E|\gamma_1\rangle+2\sum_{\gamma,\delta}\langle\gamma_1\delta_2|F_{12}|\gamma_1\delta_2\rangle-\sum_{\gamma,\delta}\langle\gamma_1\delta_2|F_{12}|\gamma_2\delta_1\rangle+N. \quad (2.9)$$

Introducing Schrödinger's representation[3] for states and operators in W, we obtain

$$W=W_1+W_2+W_3+N, \quad (2.10)$$

$$W_1=2\sum_\gamma\int\int dq_1'dq_1''\langle\gamma_1|q_1'\rangle\times\langle q_1'|E_1|q_1''\rangle\langle q_1''|\gamma_1\rangle, \quad (2.11)$$

$$W_2=2\sum_{\gamma,\delta}\int\int dq_1'dq_2''F(q_1',q_2'')\langle q_1'|\gamma_1\rangle\times\langle\gamma_1|q_1'\rangle\langle q_2''|\delta_2\rangle\langle\delta_2|q_2''\rangle, \quad (2.12)$$

$$W_3=-\sum_{\gamma,\delta}\int\int dq_1'dq_2''F(q_1'q_2'')\langle q_2''|\gamma_2\rangle\times\langle\gamma_1|q_1'\rangle\langle q_1'|\delta_1\rangle\langle\delta_2|q_2''\rangle. \quad (2.13)$$

[For expressions equivalent to (2.11)–(2.13) but not summed over spins, see reference 4 and Dirac[5].] Let

$$\langle q_2''|\rho|q_1'\rangle=\sum_{s_1'}\langle q_2''s_2''|\rho|q_1's_1'\rangle\delta_{s_1's_2''} \quad (2.14)$$

$$=\sum_{s_1',\alpha}\langle q_2''s_2''|\alpha_2\rangle\langle\alpha_1|q_1's_1'\rangle\delta_{s_1's_2''} \quad (2.15)$$

$$=2\sum_\gamma\langle q_2''|\gamma_2\rangle\langle\gamma_1|q_1'\rangle. \quad (2.16)$$

We refer to the matrix $\langle q_2''|\rho|q_1'\rangle$ as ρ the spinless density matrix. It is the trace with respect to spin of the Dirac density matrix.[5] ρ satisfies $\rho^2=2\rho$ its characteristic values are 2 or zero, and the states belonging to characteristic value 2 are the spinless states $|\gamma\rangle$. $\langle q'|\rho|q'\rangle$ is the total electron density at q'. In terms of ρ,

$$W_1=\int\int dq_1'dq_1''\langle q_1'|E|q_1''\rangle\langle q_1''|\rho|q_1'\rangle, \quad (2.17)$$

$$W_2=\tfrac{1}{2}\int\int dq_1'dq_2''F(q_1',q_2'')\times\langle q_1'|\rho|q_1'\rangle\langle q_2''|\rho|q_2''\rangle, \quad (2.18)$$

$$W_3=-\tfrac{1}{4}\int\int dq_1'dq_2''F(q_1',q_2'')\times\langle q_2''|\rho|q_1'\rangle\langle q_1'|\rho|q_2''\rangle, \quad (2.19)$$

$$=-\tfrac{1}{4}\int\int dq_1'd\xi_2'F(q_1',q_1'+\xi_2')\langle q_1'+\xi_2'|\rho|q_1'\rangle\langle q_1'|\rho|q_1'+\xi_2'\rangle. \quad (2.20)$$

[5] P. A. M. Dirac, Proc. Cambridge Phil. Soc. **26**, 376 (1930).

By taking the variation of W with respect to ρ subject to the condition $\rho^2=2\rho$ we obtain equations for the components of ρ. These equations are essentially equivalent to the Hartree-Fock equations. If we make further assumptions regarding the form of ρ, we obtain other systems of equations by variation of W subject to the conditions imposed by our assumptions. The "statistical field" treatment involves assumptions regarding the form of ρ in phase space, and so we transform ρ in W_1, W_2, W_3 to a mixed representation using

$$\langle q'|\rho|q'\rangle=\int dp'\langle q'|\rho|p'\rangle\langle p'|q'\rangle, \quad (2.21)$$

where

$$\langle p'|q'\rangle=\exp(-i\mathbf{p}'\cdot\mathbf{q}'/\hbar). \quad (2.22)$$

We let $\langle q'|\rho|p'\rangle\langle p'|q'\rangle=\rho(q',p')$. Then

$$\langle q'|\rho|q''\rangle=\int dp'\langle q'|\rho|p'\rangle\langle p'|q''\rangle=\int dp'\rho(q',p')\exp[i\mathbf{p}'\cdot(\mathbf{q}'-\mathbf{q}'')/\hbar], \quad (2.23)$$

since

$$\langle p'|q''\rangle=\langle p'|q'\rangle\exp[i\mathbf{p}\cdot(\mathbf{q}'-\mathbf{q}'')/\hbar]. \quad (2.24)$$

In terms of $\rho(q',p')$:

$$W_1=\int\int dq_1'dp_1'E(q_1',p_1')\rho(q_1',p_1'), \quad (2.25)$$

$$W_2=\tfrac{1}{2}\int\int\int\int dq_1'dq_2''dp_1'dp_2''F(q_1',q_2'')\times\rho(q_1',p_1')\rho(q_2'',p_2''), \quad (2.26)$$

$$W_3=-\tfrac{1}{4}\int\int\int\int dq_1'd\xi_2'dp_1'dp_2''\times F(q_1',q_1'+\xi_2')\times\rho(q_1',p_1')\times\bar{\rho}(q_1',p_2'')\exp[i\xi_2'\cdot(\mathbf{p}_2''-\mathbf{p}_1')/\hbar] \quad (2.27)$$

$$=-\frac{e^2\hbar^2}{4\pi}\int\int\int dq_1'dp_1'dp_2''\frac{1}{|\mathbf{p}_1'-\mathbf{p}_2''|^2}\times\rho(q_1',p_1')\bar{\rho}(q_1',p_2''). \quad (2.28)$$

[For details of the integration over ξ_2' in (2.27) to obtain (2.28) see references 4 and 5.] Let

$$\rho(q',p')=2/h^3, \quad p'\leq P(q'), \quad (2.29)$$

$$\rho(q',p')=0, \quad p'>P(q'). \quad (2.30)$$

Assume that at q' the surface $P(q')$ is spherical. This enables us to perform the integrations over p' in W_1, W_2, and W_3. With $E_1=(1/2m)p_1'^2+eV(q')$ and

$F=e^2/|\mathbf{q}_1'-\mathbf{q}_2''|$, we obtain

$$W_1=\frac{2}{h^3}\int dq_1'\left(\frac{4\pi}{10m}P_1^5+\frac{4\pi}{3}eV_1P_1^3\right), \tag{2.31}$$

$$W_2=\frac{2e^2}{h^6}\int\int dq_1'dq_2''\frac{4\pi}{3}P_1^3\frac{4\pi}{3}P_2^3\frac{1}{|\mathbf{q}_1'-\mathbf{q}_2''|}, \tag{2.32}$$

$$W_3=-\frac{2e^2}{h^6}\int dq_1'2\pi P_1^4, \tag{2.33}$$

and for the total number of electrons, E, we obtain

$$E=2\int\int dq_1'dp_1'\rho(q_1',p_1')=\frac{2}{h^3}\int dq_1'\frac{4\pi}{3}P_1^3. \tag{2.34}$$

If we define $\rho(q',p')=2/h^3$ in an infinite range in p', then

$$\langle q'|\rho|q''\rangle=(2/h^3)\int dp'$$

$$\times\exp[i\mathbf{p}'\cdot(\mathbf{q}'-\mathbf{q}'')/\hbar]=2\delta(\mathbf{q}'-\mathbf{q}''). \tag{2.35}$$

With ρ defined as a δ function in this way, the condition $\rho^2=2\rho$ will be satisfied. However, when we restrict ourselves to a finite range in p' (i.e., a finite number of particles), we obtain an approximation to the δ function which is better the larger either P or $q_1'-q_2''$. Consequently, the condition $\rho^2=2\rho$ will be only approximately satisfied. In determining our variational equations, we use (2.34) fixing the number of electrons as a constraint condition instead of $\rho^2=2\rho$. We obtain

$$(1/2m)P_1^2-(2e^2/h)P_1+e\Phi_1=\lambda, \tag{2.36}$$

where

$$\Phi_1=V_1+\int dq_2''\frac{8\pi e}{3h^3}P_2^3\frac{1}{|\mathbf{q}_2''-\mathbf{q}_1'|}. \tag{2.37}$$

λ is chosen so that (2.34) is satisfied. [Dirac[5] obtained $-4e^2/h$ for the coefficient of P_1 in (2.36); see Brillouin[6] for explanation.]

Equations (2.36), (2.37), and our condition on λ give us a uniquely determined problem when V is specified. However, implicit in our derivation is the condition $P_1\geq 0$. With this condition, no continuous solution may exist. To see this, introduce spherical coordinates around the "center of nuclear charge" and let the nuclear charges be eZ_i. Then as $r\to\infty$, $\Phi_1\to-e(\sum_i Z_i-E_r)/r$. We mean by E_r the number of electrons enclosed in a sphere of radius r. For a neutral or positively ionized system, $\Phi_1<0$ and tends to 0 as above. P_1 must tend to zero to satisfy (2.34). Therefore, λ must be zero. At sufficiently small P_1, $(1/2m)P_1^2$ is negligible with respect to $(2e^2/h)P_1$, and so $(2e^2/h)P_1\doteq e\Phi_1<0$, a contradiction. If we attempt to apply this theory to positively charged ions, there will be a surface where Φ_1 is zero, and again P_1 will have to be negative in the neighborhood of this surface.

We then consider the possibility of having a discontinuous solution by confining the charge distribution to a finite region of space. Let the distribution be contained within $C(r)$. Consider another distribution, P_1', satisfying (2.34), (2.36), (2.37) contained within $C(r)+c(r)$, where $C(r)+c(r)$ encloses $C(r)$, and v, the volume difference in the two distributions, is small. Let W' be the energy of the second distribution and let W_v' and E_v' be the energy and number of particles in v. Let W_C' be the part of W' in $C(r)$. Then

$$W'-W=W_C'-W+W_v'. \tag{2.38}$$

Inside $C(r)$ the difference $P_1'-P_1$ is a variation δP_1. The solution P_1 has been chosen so that $\delta W=\lambda\delta E$ for all small variations δP_1, within $C(r)$. Consequently,

$$W_C'-W=\lambda(E_C'-E)=-\lambda E_v'. \tag{2.39}$$

If we choose $c(r)$ so that $C(r)+c(r)$ contains a small local perturbation of $C(r)$ in which P_1 is essentially constant, then

$$W_v'=\int_v dq_1'\left(\frac{3}{10m}P_1^2-\frac{3e^2}{h}P_1+e\Phi_1\right)\frac{8\pi}{3h^3}P_1^3$$

$$\doteq\left(\frac{3}{10m}P_1^2-\frac{3e^2}{h}P_1+e\Phi_1\right)E_v'. \tag{2.40}$$

Substituting from (2.36) for λ in (2.39), we obtain from (2.38), (2.39), (2.40):

$$W'-W=\left(-\frac{1}{5m}P_1^3+\frac{e^2}{2h}P_1\right)E_v'. \tag{2.41}$$

$W'<W$ if $P_1>5e^2m/2h$, and so we take as a boundary value $P_1=5e^2m/2h$. Outside this boundary we have no charge and $P_1=0$.

The diagonal elements of the spinless density matrix are now given by $\langle q_1'|\rho|q_1'\rangle=(8\pi/3h^3)P_1^3(q_1')$. We introduce $\rho_1=(8\pi/3h^3)P_1^3$. In terms of ρ_1 and P_1,

$$eE=-e\int dq_1'\rho_1 \quad \text{(total charge)}, \tag{2.42}$$

$$K=\frac{4\pi}{5mh^3}\int dq_1'P_1^5 \quad \text{(kinetic energy)}, \tag{2.43}$$

$$V=e\int dq_1'\rho_1V_1 \quad \text{(potential energy for electron-nuclei interaction)}, \tag{2.44}$$

$$I=\frac{e^2}{2}\int\int dq_1'dq_2'\frac{\rho_1\rho_2}{|\mathbf{q}_1'-\mathbf{q}_2'|} \quad \text{(electron mutual potential interaction energy)}, \tag{2.45}$$

$$U=-\frac{4\pi e^2}{h^6}\int dq_1'P_1^4 \quad \text{(exchange energy)}, \tag{2.46}$$

[6] L. Brillouin, *L' Atomde Thomas-Fermi* (Hermann & Cie, Paris, 1934).

$$N=\sum_{i>j} e^2 Z_i Z_j / b_{ij} \quad \text{(nuclear mutual potential interaction energy)}, \tag{2.47}$$

$$\mathbf{f}_i=-\sum_{j,\,j\neq i}\left\{\frac{eZ_j}{b_{ij}^3}\mathbf{b}_{ij} - \int |d\mathbf{r}_1|\frac{e\rho_1}{r_i^3}\mathbf{r}_i\right\}, \quad \text{(force per unit charge on nucleus } i\text{)}. \tag{2.48}$$

Here $\mathbf{b}_{ij}$ is a vector from nucleus i to nucleus j, and $\mathbf{r}_i$ is the radius vector from nucleus i. We have certain relations between these integrals. In the first place, multiplying (2.36) by $(8\pi/h^3)P_1^5$, we see that the following equation must hold:

$$5K+3V+6I+4U=3\lambda E. \tag{2.49}$$

Second, we derive a general virial theorem which involves these integrals. If we have any two charges Q_1, Q_2 in Coulomb interaction located at $\mathbf{r}_1$ and $\mathbf{r}_2$, then the virial, χ, is

$$\chi=-Q_1\mathbf{f}_1\cdot\mathbf{r}_1-Q_2\mathbf{f}_2\cdot\mathbf{r}_2 = -\frac{Q_2Q_1}{r_{12}^3}\mathbf{r}_{12}\cdot(\mathbf{r}_2-\mathbf{r}_1)=-V. \tag{2.50}$$

Consequently, the virial of any Coulomb distribution will be minus the total potential interaction energy of the system. Applying this theorem to the statistical field distribution,

$$\chi=-e\sum_i Z_i\mathbf{r}_{1i}\cdot\mathbf{f}_i + e\int |d\mathbf{r}_1|\rho_1\mathbf{r}_1\cdot\nabla\Phi_1=-N-V-I. \tag{2.51}$$

From (2.36),

$$e\nabla\Phi_1=-(1/m)P_1\nabla P_1+(2e^2/h)\nabla P_1. \tag{2.52}$$

$$e\rho_1\nabla\Phi_1=-\frac{8\pi}{3h^3}\nabla\left(\frac{P_1^5}{5m}-\frac{2e^2}{4h}P_1^4\right). \tag{2.53}$$

$$e\int |d\mathbf{r}_1|\rho_1\mathbf{r}_1\cdot\nabla\Phi_1 = -\frac{8\pi}{3h^3}\int |d\mathbf{r}_1|\mathbf{r}_1\cdot\nabla\left(\frac{P_1^5}{5m}-\frac{2e^2}{h}\frac{P_1^4}{4}\right) = -\frac{8\pi}{3h^3}\sum_i\int_{s_i} d\mathbf{S}_i\cdot\mathbf{r}_1\left(\frac{P_1^5}{5m}-\frac{e^2}{2h}P_1^4\right) + \frac{8\pi}{h^3}\int |d\mathbf{r}_1|\left(\frac{P_1^5}{5m}-\frac{e^2}{2h}P_1^4\right). \tag{2.54}$$

The surface integrals are taken negatively over small spheres surrounding the nuclei, and positively over the boundary $P_1=5e^2m/2h$. The boundary value of P_1 makes the integrand of the surface integral vanish over the boundary. At a nucleus $P_1^4=O(1/r^2)$ and $rP_1^4dS=O(r)d\Omega$ so that in the limit $r\to 0$ the surface integrals at the nuclei vanish. Thus

$$e\int |d\mathbf{r}_1|\rho_1\mathbf{r}_1\cdot\nabla\Phi_1=2K+U, \tag{2.55}$$

and from (2.51) we obtain for χ_N, the virial of the nuclear charges only,

$$-\chi_N=e\sum Z_i\mathbf{r}_{1i}\cdot\mathbf{f}_i=N+V+I+2K+U=W+K. \tag{2.56}$$

Dirac[6] obtained the equations for the statistical field by obtaining a function $K(p',q')$ representing the energy at a point (p',q') in phase space, letting ρ equal $2/h^3$ in this function, and noting that $K(p',q')$ must be constant when p' equals P. From this derivation one does not obtain any boundary conditions. These must be obtained from additional physical considerations (see Brillouin[6]). The foregoing treatment of the statistical field is essentially a generalization to molecules of a method presented by Jensen.[7] Fock[8] obtained a general virial theorem for the Thomas-Fermi field. The identity (2.49) was obtained by Thomas.[9]

If we make a displacement $d\mathbf{r}_{1j}$ in nucleus j, then χ_N, K, W and the forces may be considered functions of $\mathbf{r}_{1j}$, the position of nucleus j. We have

$$-d\chi_N=-(\nabla_j\chi_N)\cdot d\mathbf{r}_{1j}=(\nabla_jW+\nabla_jK)\cdot d\mathbf{r}_{1j}. \tag{2.57}$$

As $d\mathbf{r}_{1j}$ is an arbitrary small displacement,

$$-\nabla_j\chi_N=\nabla_jW+\nabla_jK. \tag{2.58}$$

Let

$$\mathbf{F}_j=eZ_j\mathbf{f}_j. \tag{2.59}$$

$\mathbf{F}_j$ is the total force acting on nucleus j. Then, as work $\mathbf{F}_j\cdot d\mathbf{r}_{1j}$ is done at the expense of the energy of the system.

$$\mathbf{F}_j=-\nabla_jW. \tag{2.60}$$

Then

$$-\nabla_j\chi_N=\nabla_j(\textstyle\sum_i\mathbf{F}_i\cdot\mathbf{r}_{1i}) =\mathbf{F}_j+\sum_i(\mathbf{r}_{1i}\cdot\nabla_j)\mathbf{F}_i+\sum_i[\mathbf{r}_{1i}\times(\nabla_j\times\mathbf{F}_i)] =-\mathbf{F}_j+\nabla_jK. \tag{2.61}$$

We have an equation of form (2.61) for each nucleus, which, together with $\sum\mathbf{F}_i=0$, serve to completely determine the forces as a function of position once the total kinetic energy is known as a function of nuclear position. Then the energy W may be determined from (2.60) so that once the kinetic energy is known as a function of nuclear position, the energy and forces may be regarded as determined.

For a neutral diatomic molecule, Eqs. (2.61) are particularly simple. If we consider displacements along

[7] H. Jensen, Z. Physik **93**, 232 (1935).
[8] V. Fock, Physik. Z. Sowjetunion **1**, 747 (1932).
[9] L. H. Thomas, J. Chem. Phys. **22**, 1758 (1954).

the common nuclear axis, we obtain

$$\frac{dF}{da}+\frac{2F}{a}=\frac{1}{2a}\frac{dK}{da}. \tag{2.62}$$

a is the internuclear half-distance. $F=0$ and $K=K_0$ when $2a=2a_0$, where $2a_0$ is the distance between nuclei when the boundaries $P=5e^2m/2h$ for the isolated atoms are just touching. K_0 is the kinetic energy of the two isolated atoms. Solving Eq. (2.62) subject to its boundary conditions, we obtain

$$F=\frac{K-K_0}{2a}+\frac{1}{2a^2}\int_a^\infty d\xi(K-K_0). \tag{2.63}$$

As $a\to 0$, $F\to Z_1Z_2/4a^2$, and so

$$\int_0^\infty d\xi(K-K_0)=Z_1Z_2/2. \tag{2.64}$$

Our integral equation (2.36) with its boundary conditions can in general be reduced to a differential equation with boundary conditions. If $\nabla^2\Phi_1=-4\pi\rho_1$, then $\Phi_1(\mathbf{r}_1)=\int|d\mathbf{r}_2|\rho(\mathbf{r}_2)/|\mathbf{r}_1-\mathbf{r}_2|$, whence we deduce from (2.37) and (2.36):

$$-\nabla^2e\Phi_1=\nabla^2\left(\frac{1}{2m}P_1^{\frac{3}{2}}-\frac{2e^2}{h}P_1-\lambda\right)$$

$$=\frac{32\pi^2e^2}{3h^3}P_1^{\frac{3}{2}},\quad P_1\geq\frac{5e^2m}{2h}, \tag{2.65}$$

$$-\nabla^2e\Phi_1=0,\quad P_1<\frac{5e^2m}{2h}. \tag{2.66}$$

The boundary conditions are:

$$\Phi_1\to-Z_ie/r_i\quad\text{as}\quad r_i\to 0, \tag{2.67}$$

$$\Phi_1\to-(\textstyle\sum Z_i-E)e/r\quad\text{as}\quad r\to\infty. \tag{2.68}$$

λ is chosen so that

$$-\frac{1}{4\pi e}\int|d\mathbf{r}|\nabla^2\Phi_1=-\frac{1}{4\pi e}\int_S d\mathbf{S}\cdot\nabla\Phi_1=E-\textstyle\sum Z_i. \tag{2.69}$$

S in (2.69) is any surface outside the boundary. It is convenient to introduce a new variable $\bar{V}$ defined by

$$\bar{V}=\frac{1}{2m}\left(P_1-\frac{2e^2m}{h}\right)^2, \tag{2.70}$$

so that $-e\Phi_1=\bar{V}-(2e^4m/h^2)\lambda$. Then in terms of $\bar{V}$, (2.65) to (2.69) become

$$\nabla^2\bar{V}=\frac{32\pi^2e^2}{3h^3}\left[\frac{2e^2m}{h}+(2m\bar{V})^{\frac{1}{2}}\right]^3,\quad \bar{V}\geq\frac{e^4m}{8h^2}, \tag{2.71}$$

$$\nabla^2\bar{V}=0,\quad \bar{V}<\frac{e^4m}{8h^2}, \tag{2.72}$$

$$\bar{V}\to Z_ie^2/r_i\ \text{as}\ r_i\to 0, \tag{2.73}$$

$$\bar{V}\to e^2(\textstyle\sum Z_i-E)/r+2e^4m/h^2+\lambda\ \text{as}\ r\to\infty. \tag{2.74}$$

λ is chosen so that

$$\frac{1}{4\pi e^2}\oint d\mathbf{S}\cdot\nabla\bar{V}=E-\textstyle\sum Z_i. \tag{2.75}$$

If we choose a value for λ, then we have a perfectly definite elliptic boundary value problem in (2.71), (2.72), (2.73), and (2.74). Obtaining a solution for this λ we test to see if (2.75) is satisfied, and if not obtain new solutions with other values of λ until we satisfy (2.75).

For a neutral system we can actually evaluate λ. From Eq. (2.36), the boundary $P_1=5e^2m/2h$ is an equipotential and so the flux through it must be everywhere positive or everywhere negative. Then from (2.69) with right-hand size zero, $\nabla\Phi$ must be zero. Then as Φ is constant, $\nabla\Phi$ zero, and no charge is outside the boundary, Φ must be constant everywhere outside. Then from (2.68) Φ must be zero on and outside the boundary. Thence from (2.36) we obtain

$$\lambda=-15e^4m/8h^2, \tag{2.76}$$

and for a neutral system equations (2.72) and (2.74) may be replaced by the condition

$$\bar{V}=e^4m/8h^2\ \text{on and outside boundary}\quad P_1=5e^2m/2h. \tag{2.77}$$

3. NUMERICAL SOLUTION OF STATISTICAL FIELD EQUATIONS FOR HOMONUCLEAR DIATOMIC MOLECULES

We wish to replace our system of equations (2.71)–(2.75) by finite difference approximations and solve the resulting algebraic equations, for the case of a homonuclear diatomic molecule. For this purpose it is convenient to introduce a coordinate system which makes it easy to represent the boundary conditions and the integrals by finite difference approximations. It is also advantageous to choose coordinates which are such that frequent changes of interval are not required in order to get a network suitable for solving the problem. Here we choose the same system used by Hund,[1] except we take an exponential scale in the variable ξ. A coordinate transformation is introduced so that all distances are measured in terms of the nuclear half-distance, a. Then in the azimuthal plane (see Fig. 1) we define coordinates (η,t):

$$1/\xi=1/r_1+1/r_2, \tag{3.1}$$

$$\xi=e^t, \tag{3.2}$$

$$\eta=\cos\theta_1+\cos\theta_2, \tag{3.3}$$

$$\xi'=a\xi. \tag{3.4}$$

Let g_{ij} be the covariant components of the metric tensor associated with the coordinate system (3.1), (3.3). (See Hund[1] for useful properties of g_{ij}.) Let the

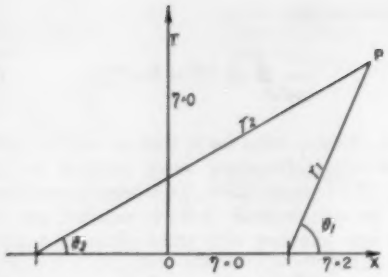


FIG. 1. The azimuthal plane. Quantities shown are used to define coordinates for the numerical solution of the statistical field equations.

boundary where $P_1=5e^2m/2h$ be

$$t_b=t_b(\eta). \tag{3.5}$$

We introduce atomic units in the measurement of all physical quantities. In these units we have:

$$\rho_1=(1/3\pi^2)P_1, \tag{3.6}$$

$$E=4\pi\int_0^2\int_{-\infty}^{t_b} dtd\eta\rho_1\xi'^3 g_{11}, \tag{3.7}$$

$$K=\frac{6\pi}{5}\int_0^2\int_{-\infty}^{t_b} dtd\eta\rho_1 P_1^2\xi'^3 g_{11}, \tag{3.8}$$

$$U=-3\int_0^2\int_{-\infty}^{t_b} dtd\eta\rho_1 P_1\xi'^3 g_{11}, \tag{3.9}$$

$$V=-4\pi Z\int_0^2\int_{-\infty}^{t_b} dtd\eta\rho_1 P_1\xi'^2 g_{11}. \tag{3.10}$$

In obtaining the differential system (2.71)–(2.74) in these coordinates it is convenient to introduce a new dependent variable u, defined by

$$\bar{V}=(1/2\pi^2)(u^2/\xi'). \tag{3.11}$$

Then we have

$$\frac{\partial^2u^2}{\partial t^2}-\frac{\partial u^2}{\partial t}+\left(\frac{\partial}{\partial\eta}-\frac{r^2}{\xi^2}\right)\frac{\partial u^2}{\partial\eta}+\frac{r^2}{\xi^2}\frac{\partial^2u^2}{\partial\eta^2}$$

$$=\frac{8}{3\pi}g_{11}\xi'^3(1+u\xi'^{-1})^3, \quad t\le t_b(\eta) \tag{3.12}$$

$$=0, \quad t>t_b(\eta). \tag{3.13}$$

$$u\to(2Z)^{\frac{1}{2}}\pi \quad \text{as} \quad t\to-\infty, \tag{3.14}$$

$$u\to 0 \quad \text{as} \quad t\to+\infty. \tag{3.15}$$

Along the nuclear axis, $r=0$, g_{22} is singular and at the nodal point, $r=x=0$, g_{11} and g_{22} are both singular. Along the nuclear bisector $x=0$ we have symmetry. At these places we obtain the following limiting forms for (3.12):

$$\frac{\partial^2u^2}{\partial t^2}-\frac{\partial u^2}{\partial t}+\left(\frac{\partial}{\partial\eta}-\frac{r^2}{\xi^2}\right)\frac{\partial u^2}{\partial\eta}$$

$$=\frac{8}{3\pi^2}g_{11}\xi'^3(1+u\xi'^{-1})^3, \quad r=0, \quad x\neq 0, \tag{3.16}$$

$$\left(\frac{\partial u^2}{\partial t}\right)_+-\left(\frac{\partial u^2}{\partial t}\right)_-$$

$$=\frac{8}{3\pi^2}\xi'^3(1+u\xi'^{-1})^3, \quad r=0, \quad x=0, \tag{3.17}$$

$$\frac{\partial^2u^2}{\partial t^2}-\frac{\partial u^2}{\partial t}+\frac{r^2}{\xi^2}\frac{\partial^2u^2}{\partial\eta^2}$$

$$=\frac{8}{3\pi^2}g_{11}\xi'^3(1+u\xi'^{-1})^3, \quad r\neq 0, \quad x=0. \tag{3.18}$$

In (3.17), by $(\partial u^2/\partial t)_+$ we mean the limit of this derivative as we approach 0 along $x=0$ (i.e., $\xi>\frac{1}{2}$) and by $(\partial u^2/\partial t)_-$ the limit as we approach along $r=0$ (i.e., $\xi<\frac{1}{2}$). We obtain this limiting form by looking at the power series development of $\bar{V}$ about 0:

$$\bar{V}_+=\bar{V}_0+\left(\frac{\partial\bar{V}}{\partial\xi}\right)_+(\xi-\tfrac{1}{2})+\frac{1}{2!}\left(\frac{\partial^2\bar{V}}{\partial\xi^2}\right)_+$$

$$\times(\xi-\tfrac{1}{2})^2+\frac{1}{2!}\frac{\partial^2\bar{V}}{\partial\eta^2}\eta^2+\cdots, \tag{3.19}$$

$$\bar{V}_-=\bar{V}_0+\left(\frac{\partial\bar{V}}{\partial\xi}\right)_-(\xi-\tfrac{1}{2})+\frac{1}{2!}\left(\frac{\partial^2\bar{V}}{\partial\xi^2}\right)_-$$

$$\times(\xi-\tfrac{1}{2})^2+\frac{1}{2!}\frac{\partial^2\bar{V}}{\partial\eta^2}\eta^2+\cdots. \tag{3.20}$$

Terms in η and $\eta\xi$ do not occur on account of symmetry. We may show that in the neighborhood of $\eta=0$, $\xi=\frac{1}{2}$:

$$\xi=\tfrac{1}{2}(1-x^2+\tfrac{1}{2}y^2+\tfrac{1}{2}z^2)+O(x^4,x^2y^2,\text{etc.}), \tag{3.21}$$

$$\eta=2xy^2+O(x^2y^3,x^4y,\text{etc.}), \tag{3.22}$$

and

$$\bar{V}_+=\bar{V}_0+\left(\frac{\partial\bar{V}}{\partial\xi}\right)_+\left(-\frac{x^2}{2}+\tfrac{1}{4}y^2+\tfrac{1}{4}z^2\right)$$

$$+O(x^4,x^2y^2,\text{etc.}), \tag{3.23}$$

$$\bar{V}_-=\bar{V}_0+\left(\frac{\partial\bar{V}}{\partial\xi}\right)_-\left(-\frac{x^2}{2}+\tfrac{1}{4}y^2+\tfrac{1}{4}z^2\right)$$

$$+O(x^4,x^2y^2,\text{etc.}). \tag{3.24}$$

Then

$$\partial^2\bar{V}/\partial x^2 = -(\partial\bar{V}/\partial\xi)_-, \tag{3.25}$$

$$\partial^2\bar{V}/\partial y^2 = \partial^2\bar{V}/\partial Z^2 = \tfrac{1}{2}(\partial\bar{V}/\partial\xi)_+, \tag{3.26}$$

$$\nabla^2\bar{V} = (\partial\bar{V}/\partial\xi)_+ - (\partial\bar{V}/\partial\xi)_-. \tag{3.27}$$

Making changes of coordinates and dependent variable, we obtain (3.17) from (3.27), (3.11), and (2.71).

We now replace Eqs. (3.12), (3.16), (3.17), and (3.18) by difference equations over a network. We choose I intervals in η of size $\Delta\eta = 2/I$. We choose K intervals in t of size $(\Delta t)_{k+\frac{1}{2}}$ where $k = 0, 1, \cdots, K$. Values of t, t_k, are so chosen that $\xi = \frac{1}{2}$ (i.e., the point O) is in the set. Call this value of t, t_n. A point in the network belonging to $\eta_i = i\Delta\eta$, t_k is labeled (i,k). Then derivatives are replaced by finite difference approximations as follows:

$$\left(\frac{\partial^2 u^2}{\partial t^2}\right)_{i,k} = \frac{2}{(\Delta t)_{k-\frac{1}{2}} + (\Delta t)_{k+\frac{1}{2}}}\left[\frac{u_{i,k+1}{}^2 - u_{i,k}{}^2}{(\Delta t)_{k+\frac{1}{2}}} - \frac{u_{i,k}{}^2 - u_{i,k-1}{}^2}{(\Delta t)_{k-\frac{1}{2}}}\right], \quad [\text{except for } (0,n)], \tag{3.28}$$

$$\left(\frac{\partial u^2}{\partial t}\right)_{i,k} = \frac{1}{2}\left[\frac{u_{i,k+1}{}^2 + u_{i,k}{}^2}{(\Delta t)_{k+\frac{1}{2}}} - \frac{u_{i,k}{}^2 + u_{i,k-1}{}^2}{(\Delta t)_{k-\frac{1}{2}}}\right], \quad [\text{except for } (0,n)], \tag{3.29}$$

$$\left(\frac{\partial^2 u^2}{\partial \eta^2}\right)_{i,k} = 2\left[\frac{u_{i+1,k}{}^2 - u_{i,k}{}^2}{(\Delta\eta)^2}\right], \quad i = 0, \quad k > n, \tag{3.30}$$

$$\left(\frac{\partial^2 u^2}{\partial \eta^2}\right)_{i,k} = \frac{u_{i+1,k}{}^2 - 2u_{i,k}{}^2 + u_{i-1,k}{}^2}{(\Delta\eta)^2}, \quad 0 < i < I, \tag{3.31}$$

$$\left(\frac{\partial u^2}{\partial \eta}\right)_{i,k} = \frac{u_{i+1,k}{}^2 - u_{i,k}{}^2}{\Delta\eta}, \quad i = 0, \tag{3.32}$$

$$\left(\frac{\partial u^2}{\partial \eta}\right)_{i,k} = \frac{u_{i+1,k}{}^2 - u_{i-1,k}{}^2}{2\Delta\eta}, \quad 0 < i < I, \tag{3.33}$$

$$\left(\frac{\partial u^2}{\partial \eta}\right)_{i,k} = \frac{u_{i,k}{}^2 - u_{i-1,k}{}^2}{\Delta\eta}, \quad i = I. \tag{3.34}$$

By means of (3.28)–(3.34) the left-hand sides of each of equations (3.12), (3.16), (3.18) may be represented by choosing the appropriate approximation. The left-hand side of Eq. (3.17) may be represented by (3.28) if at this point we take

$$g_{11} = 2/[(\Delta t)_{k+\frac{1}{2}} + (\Delta t)_{k-\frac{1}{2}}], \tag{3.35}$$

and then multiply the right-hand side of (3.17) by this g_{11}. Let us call the left-hand side of any one of Eqs. (3.12), (3.16)–(3.18), C_{ik}, the value of i and k being sufficient to identify which form for the left-hand side is appropriate. Then for a neutral molecule the system of equations we wish to solve is

$$u_{i,k} = \bar{u}_{i,k} \quad \text{when} \quad k = 0, \tag{3.36}$$

$$f_{i,k} = C_{i,k} - \frac{8}{3\pi^2} g_{11,i,k}\xi_k{}'^3(1 + u_{i,k}\xi_k{}'^{-1})^3, \quad k = 1,2,\cdots k_b(i), \tag{3.37}$$

$$u_{i,k} = \xi_k{}'^4/4, \quad k > k_b(i). \tag{3.38}$$

In (3.36), we have replaced the boundary condition (3.14) by the condition that the solution is prescribed on a coordinate line belonging to t_0. The value of t_0 is chosen so that on this line we may prescribe as boundary values, values taken from the solution of the statistical field equations for the neutral atom. The criterion chosen for this was that, assuming no screening by the electronic charge, the effect of the field of the other nucleus should not affect the kinetic energy contained within this coordinate line by as much as 0.1 atomic unit. This is a very safe criterion since, unless the nuclei are very close, there is considerable screening.

For Eqs. (3.28)–(3.34) one may readily determine that the local truncation error in approximating derivatives by difference quotients is of order $(\Delta t)^2$ or $(\Delta\eta)^2$ at all points in the interior of our region ($i \neq 0$, $i \neq I$). Along limiting lines, the error is proportional to $\Delta\eta$ and at the nodal point, O, to Δt.

However, by examining the power series development of the solution along limiting lines and at the nodal point, we can show that the truncation error is really proportional to $(\Delta x)^2$ and $(\Delta r)^2$ where Δx and Δr are distances between adjacent points in the original x, r coordinates. Hence in terms of actual distances between adjacent net-points our approximations for derivatives are of second order throughout.

The integrals (3.7)–(3.10) may be expressed in terms of $C_{i,k}$. Since there are truncation errors affecting the second differences of $C_{i,k}$, there is no advantage in making use of an integration formula which takes account of differences of the integrand higher than the second difference. Hence we use the trapezoidal rule, which just accounts for second differences, in evaluating integrals. The integrals (3.7)–(3.10) are approximated by:

$$P_{i,k}{}' = 1 + u_{i,k}\xi_k{}'^{-1}, \tag{3.39}$$

$$E = \frac{\Delta\eta}{2\pi^2}\sum_{i,k}\frac{(\Delta t)_{k-\frac{1}{2}} + (\Delta t)_{k+\frac{1}{2}}}{2} W_{i,k}C_{i,k} + E_1, \tag{3.40}$$

$$K = \frac{3\Delta\eta}{20\pi^4}\sum_{i,k}\frac{(\Delta t)_{k-\frac{1}{2}} + (\Delta t)_{k+\frac{1}{2}}}{2} \times W_{i,k}C_{i,k}P_{i,k}{}'^3 + K_1, \tag{3.41}$$

$$U = -\frac{3\Delta\eta}{8\pi^4}\sum_{i,k}\frac{(\Delta t)_{k-\frac{1}{2}} + (\Delta t)_{k+\frac{1}{2}}}{2} \times W_{i,k}C_{i,k}P_{i,k}{}' + U_1, \tag{3.42}$$

$$V = -\frac{Z\Delta\eta}{2\pi^2}\sum_{i,k}\frac{(\Delta t)_{k-\frac{1}{2}} + (\Delta t)_{k+\frac{1}{2}}}{2}\frac{W_{i,k}C_{i,k}}{\xi_k{}'} + V_1, \tag{3.43}$$

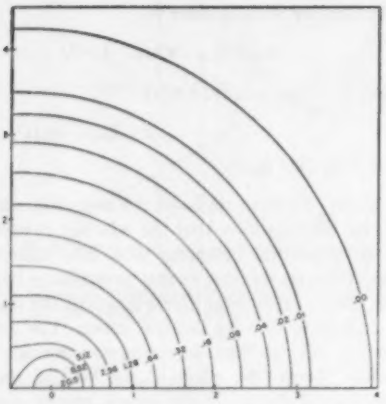


FIG. 2. Statistical field for nitrogen, $a=0.500$. Distances are in atomic units. Level curves are minus the potential $(-\Phi)$ in atomic units.

where $W_{i,k}$ is 1 in the interior of our region, $\frac{1}{2}$ along the limiting lines, and $\frac{1}{4}$ at the intersection of the coordinate line t_0 with limiting lines. E_1, K_1, U_1, V_1 are the contributions to the integrals from the field inside t_0.

The largest truncation error occurs at the boundary $t=t_b(\eta)$. If we knew the location of the boundary, we could obtain appropriate approximations, but as this boundary is initially unknown, we expect an error of first order in Δt and $\Delta\eta$ when we take differences across it, since the solution does not possess a continuous second derivative across the boundary. A method for correcting in part for this error is discussed subsequently.

The most rapidly convergent point-by-point linear iteration method currently available for LaPlace's

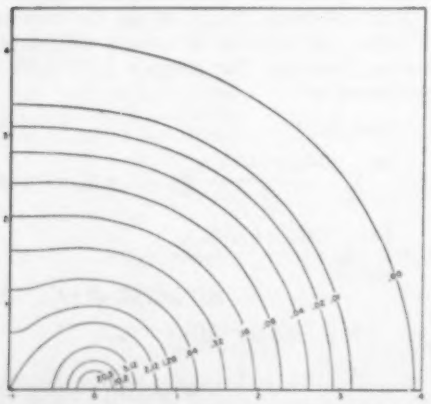


FIG. 3. Statistical field for nitrogen, $a=1.038$. Distances are in atomic units. Level curves are minus the potential $(-\Phi)$ in atomic units.

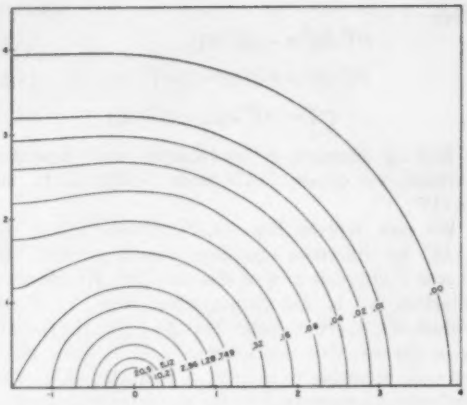


FIG. 4. Statistical field for nitrogen, $a=1.500$. Distances are in atomic units. Level curves are minus the potential $(-\Phi)$ in atomic units.

equation is the "extrapolated-Liebmann" method.[10-12] This method is here generalized to apply to our nonlinear system. Suppose we have approximate values $u_{i,k}^{(n)}$ for $u_{i,k}$ over a network. Then $f_{i,k}$ [Eq. (3.37)] is not zero but has a value $f_{i,k}^{(n)}$. Then determine $\Delta u_{i,k}^{(n)}$ and $u_{i,k}^{(n+1)}$ from

$$f_{i,k}^{(n)}+\frac{1}{\alpha}\frac{\partial f_{i,k}^{(n)}}{\partial u_{i,k}}\Delta u_{i,k}^{(n)}=0, \tag{3.44}$$

$$u_{i,k}^{(n+1)}=u_{i,k}^{(n)}+\Delta u_{i,k}^{(n)}. \tag{3.45}$$

After $u_{i,k}^{(n+1)}$ has been computed, it is substituted into $f_{i,k}^{(n)}$ for the next point, i.e., "improved values" and used wherever possible.

Let S be a function of the $u_{i,k}$ such that

$$\partial S/\partial u_{i,k}=I_{i,k}f_{i,k}. \tag{3.46}$$

$I_{i,k}$ is an "integrating factor." We compute the approximate change in S, $\delta S_{i,k}$ when $u_{i,k}^{(n+1)}$ replaces $u_{i,k}^{(n)}$ as follows:

$$\delta S_{i,k}=\frac{\partial S}{\partial u_{i,k}}\Delta u_{i,k}+\frac{1}{2}\frac{\partial^2 S}{\partial u_{i,k}^2}(\Delta u_{i,k})^2 \tag{3.47}$$

$$=\left(\frac{1}{2}-\frac{1}{\alpha}\right)\frac{\partial^2 S}{\partial u_{i,k}^2}(\Delta u_{i,k})^2, \tag{3.48}$$

and so

$$\delta S=\sum_{i,k}\delta S_{i,k}=\left(\frac{1}{2}-\frac{1}{\alpha}\right)\sum_{i,k}\frac{\partial^2 S}{\partial u_{i,k}^2}(\Delta u_{i,k})^2. \tag{3.49}$$

Equations (2.71), (2.72) can be obtained by the variation of an integral and, of course, so can (3.12), (3.13). Our difference system could be obtained also by varying a sum if we had not differentiated some of

[10] D. Young, thesis, Harvard University, 1950 (unpublished).
[11] S. P. Frankel, Mathematical Tables and Other Aids to Computation **4**, 65 (1950).
[12] J. Sheldon and L. H. Thomas, J. Appl. Phys. **24**, 235 (1953).

the terms and cancelled out powers of e. If we let

$$I_{i,k}=-[(\Delta t)_{k-\frac{1}{2}}+(\Delta t)_{k+\frac{1}{2}}]e^{-t_k}, \tag{3.50}$$

then a function S can be constructed such that (3.37) is satisfied to within terms of the same order as the truncation error. This function has an absolute minimum at the solution, whence we may show from (3.49) that $\delta S<0$ if $0<\alpha<2$ proving convergence for the interation process (3.44), (3.45).

If we perform a number of interations starting with the same iterant $u_{i,k}^{(n)}$, then δS may be considered a function of α, having a minimum at some value of α, say α'. Then there exists a value of α, α'', which is such that if we iterate with $\alpha>\alpha''$, α' decreases and if we iterate with $\alpha<\alpha''$, α' increases. We find that α'' is sharply determined and stable. It is closely the so-called "optimum value" of α which makes all error eigenfunctions decay at the same rate.[10,11]

The rate of convergence with $\alpha=\alpha''$ can be determined for LaPlace's equation in spherical coordinates with exponential spacing in r. These coordinates roughly simulate the ones we are using so far as fundamental modes are concerned, so that it is possible to make a rough estimate of the number of interactions required.

In order to solve our Eqs. (3.36)–(3.38), we assume that (3.37) applies for all k except $k=0$, $k=K$. Then if $u_{i,k}^{(n+1)}$ determined in this way is less than $\xi_k'^{\frac{1}{2}}/4$, we set $u_{i,k}^{(n+1)}$ equal to $\xi_k'^{\frac{1}{2}}/4$. This process converges and determines the boundary $t_k(i)$. We obtain convergence because, if $u_{i,k}^{(n+1)}$ [from (3.37)] is less than $\xi_k'^{\frac{1}{2}}/4$ we obtain a greater decrease in S by putting $u_{i,k}^{(n+1)}$ equal to $\xi_k'^{\frac{1}{2}}/4$. S must always decrease or remain stationary.

As mentioned previously, $t_k(\eta)$ is only roughly determined. In determining the integrals, we may correct for this error in large part. The kinetic and exchange energy densities are small near this boundary and need not be corrected. We wish to correct V. We know the value that E should have (namely $2Z$). Then when we obtain the value of E by numerical integration, E', we assume that charge $E-E'$ should be distributed uniformly about the boundary $t_k(i)$. Once this is done, the contribution of this charge to V is readily computed. For example, if the boundary is a coordinate line belonging to t_k, then

$$\Delta V=-(Z/\xi_k')\Delta E. \tag{3.51}$$

As a guide to the choice of intervals, we have available, Thomas's[9] integrations of atomic fields and Hund's[1] approximate solution for the nitrogen molecule. 14 equal steps in η and approximately 60 steps in t were taken for each case. The steps in t were chosen so that we start with a "t" corresponding to a coordinate line which circumscribes the neutral atoms and then we choose: 10 steps in t with $\Delta t=1/56$, 10 steps in t with $\Delta t=1/28$, 10 steps in t with $\Delta t=1/14$, remaining steps in t with $\Delta t=1/7$. The steps are chosen so that the nodal point falls in the network.

For further details regarding the scaling and preparation of a computer to obtain numerical results, see Ladd and Sheldon.[12] A solution for one internuclear distance is obtained in approximately 80 minutes on an International Business Machines Company Electronic Data Processing Machine (Type 701 and Associated Equipment), and requires approximately 60 iterations.

4. RESULTS FOR NITROGEN

In Figs. 2, 3, 4, and 5, we show level curves in minus the potential $(-\Phi)$ for a equal to 0.5, 1.038, 1.5, and 3.0. For $a=3.0$ the solution is spherically symmetric over almost the entire field, i.e., until we get quite close to the nuclear bisector. For $a=3.0$ and $=1.5$, the distance from the origin to a level curve, measured along positive x, is the same as the corresponding distance for a neutral atom solution, to within less

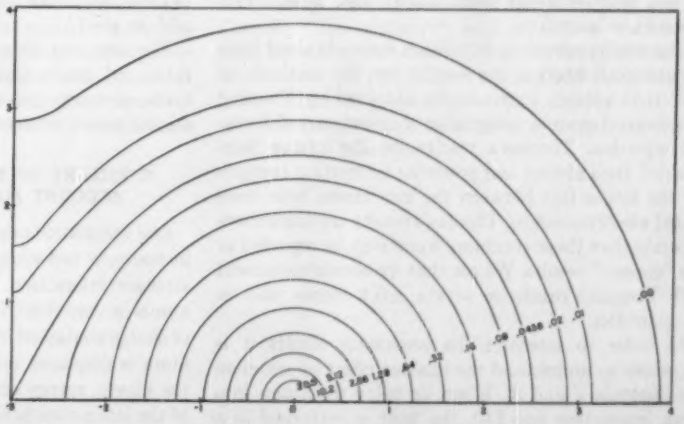


FIG. 3. Statistical field for nitrogen, $a=3.000$. Distances are in atomic units. Level curves are minus the potential $(-\Phi)$ in atomic units.

[12] D. W. Ladd and J. W. Sheldon, Proceedings of the Association for Computing Machinery, September, 1952.

TABLE I. Results of statistical field calculation for nitrogen.

1	2	3	4	5	6	7
a	0.5	1.038	1.5	3.0	8.0	8.0
$-E$	13.99	13.89	13.95	13.95	13.95	14.00
U	13.70	12.72	12.60	12.47	12.45	12.46
K	166.73	157.14	156.83	156.04	155.82	155.90
$-V$	444.68	400.02	387.93	371.35	360.93	360.90
N	49.00	23.60	16.33	8.17	3.06	3.06
I	92.20	77.20	71.33	63.62	58.58	58.51
$-W$	150.39	154.80	156.04	155.99	155.92	155.89
$-\chi_N$	16.34	2.34	0.79	0.05	0.10	0.01
$-\chi_N'$	15.69	2.04	1.33	0.23	0	0
ΔW	5.53	1.12	−0.12	−0.08	0	···
$\Delta W'$	6.06	1.08	0.62	0.03	0	···
ΔK	10.91	1.32	1.01	0.22	0	···

than 1% error. As the nuclei are moved closer together, the boundary $\Phi=0$ bows outward in the neighborhood of the nuclear perpendicular bisector. If we compare the solution for $a=1.038$ with Hund's work, we find that our solution drops off more rapidly as we leave the nucleus and also that the slope of the level curve through the nodal point is different. Of course, the lines of constant Φ are also lines of constant charge, exchange and kinetic energy density.

In Table I, we give various quantities of physical interest. E, U, and K are the charge, exchange energy, and kinetic energy obtained directly from numerical integration of our solutions. V, the potential energy of the electron-nuclear interaction, is obtained by numerical integration, but corrected for "mislocation of the boundary" (see Sec. 3). N is the computed nuclear interaction energy, I the electron mutual interaction energy obtained from E, U, K and V with the identity (2.49) and W is the total energy obtained by adding the previous energies. χ_N is the nuclear virial computed from (2.56) and χ_N' is the nuclear virial obtained by numerical integration of Eq. (2.63). ΔW is W minus $W(a=8)$ from column 6. $\Delta W'$ is the change in energy obtained by numerical integration of the force obtained from (2.63). ΔK is K minus $K(a=8)$, column 6.

The results tabulated in column 6 are obtained from a numerical solution for $a=8.0$ by the methods of Sec. 3. In column 7 are results obtained by Thomas[9] for a neutral atom by integration of an ordinary differential equation. Thomas's results for $Z=7$ have been doubled (two atoms) and potential interaction energies for the interaction between the two atoms have been added where necessary. Thomas's results are much more accurate than those of column 6 and may be regarded as the "correct" results. We see that we obtain agreement with Thomas's results to within ± 0.1 atomic unit in all quantities.

In order to interpret the remaining results it is necessary to understand the relative effect of errors on the integrals V and K. When we bring the atoms into weak interaction ($a=3.0$), the field is perturbed in a region somewhat larger than the region in which two spheres representing the neutral atoms overlap. The amount of electron-nucleus interaction energy contained in this perturbed region (about 5 atomic units) is an order of magnitude greater than the kinetic energy in this same region. Then errors in the decimals place in V may be associated with errors in the hundreths place in K. For $a=3.0$, we see that K increases by 0.22 atomic units. Then from (2.63), assuming that the kinetic energy increases linearly as the atoms are brought together, we compute the force of repulsion and then the work done. We find in this way that the energy should increase 0.03 atomic unit. The change in energy, as computed by summing K, U, V, I and N and subtracting the energy for $a=8.0$ shows a decrease of 0.08 atomic units. However ΔW computed in this way is not so accurate as $\Delta W'$ because it is sensitive to errors in V. An error of 0.322 in V for $a=3.0$ will account for the difference between ΔW and $\Delta W'$. An error of this size in V will give rise to a negligible error in K and so $\Delta W'$ is to be regarded as the more accurate value for the change in energy. Similar remarks apply to the solution for $a=1.5$. Consequently we have no real inconsistency in the results, which clearly show that the two nuclei repel one another at each separation considered. Furthermore there are no indications that we might have missed a change in the direction of the force, since it seems perfectly clear that the kinetic energy increases monotonically and hence, from (2.63) the force must be positive.

For the solution $a=0.5$, the lines of constant Φ are nearly coordinate lines of constant t. This is also true for the solution $a=8.0$. Consequently we expect the truncation error to be about the same for $a=8.0$ and $a=0.5$ so that the integrals for $a=0.5$ should be accurate to ± 0.1 atomic units. For $a=3.0$ and $a=1.5$ the lines of constant Φ cut the lines of constant t sharply in the region of interaction, and the truncation error will be larger. For best values of χ_N and ΔW, we take (from Table I) χ_N, $\Delta W'$ for $a=3.0$ and $a=1.5$; $(\chi_N+\chi_N')/2$, $(\Delta W+\Delta W')/2$ for $a=1.038$; and χ_N, ΔW for $a=0.5$.

We note that changes in exchange energy are small compared to changes in kinetic energy, so that the exchange energy plays an unimportant role in determining the forces between the nuclei.

5. FAILURE OF THE STATISTICAL FIELD TO ACCOUNT FOR MOLECULAR BINDING

Our calculations clearly show that the kinetic energy increases as two nitrogen atoms move into stronger and stronger interaction. It is of interest to consider two atoms in very weak interaction (Fig. 6). If the amount of charge contained in region "1" for an isolated neutral atom is displaced into region "2," then in first order the kinetic energy remains stationary. But if the effect of the interaction is to increase the mean charge density in "1," displacing a smaller amount of charge into

"2," then we may show that the kinetic energy increases. Let P_1' be the mean value of P_1 in region "1" with volume τ_1 for an isolated atom and let P_1'' be this mean value after the atoms are in interaction. Let $\alpha\tau_1$ be the volume of region "2" with α less than one. In region "2" the mean value of P_1 is very closely the boundary value, so let P_1 be $5/4\pi$ in region "2." Then since charge is conserved we have, approximately,

$$[P_1''^{3/2}+(5/4\pi)^{3/2}\alpha]\tau_1=2P_1'^{3/2}\tau_1. \tag{5.1}$$

Let P_1' equal $\beta(5/4\pi)$ where β must be greater than 1. Let K' and K'' be kinetic energies before and after interaction, respectively. Then

$$K''-K'=\tau_1\frac{1}{10\pi^2}\left(\frac{5}{4\pi}\right)^{5/3}[(2\beta^{3/2}-\alpha)^{5/3}-(2-\alpha)], \tag{5.2}$$

$$2\beta^{3/2}-\alpha>2-\alpha>1, \tag{5.3}$$

and so K'' is greater than K'.

For our solution for $a=3.0$ the amount of charge contained in region "2" is less than 10% that contained in region "1" for an isolated atom. Consequently, we expect that α will be small for a very weak interaction. The above arguments are readily generalized to two hetero-nuclear atoms in weak interaction and then to any number of atoms in weak interaction. Hence we do not expect that any system of atoms in weak interaction will be bound, according to the statistical field theory.

In the derivation of the equations for the statistical field we assume that at q', the surface $P(q')$ is spherical (see Sec. 2). If we eliminate this assumption and leave the surface $P(q')$ free, we may show that the requirement of minimum energy will lead us to a spherical surface $P(q')$ so that from energy considerations alone we may not improve the theory.

However, just this assumption of form regarding $P(q')$ is an important factor in accounting for the lack of success of this theory to account for molecular binding. If the surface $P(q')$ were ellipsoidal near the boundary for the isolated atom and then became more nearly spherical when the atoms come into interaction, it is possible to show for the statistical field that this would decrease the kinetic energy so that we could have binding.

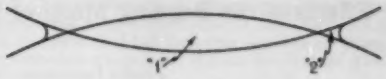


FIG. 6. Two atoms in weak interaction.

This is in fact what happens in actual molecule formation. The atoms which form molecules do not have all states of angular momentum in their outer shell and consequently we would expect $P(q')$ to be ellipsoidal. When the atoms come into interaction, new states are formed which represent a fuller complement of angular momentum states in the outer shell so that $P(q')$ will be more nearly spherical.

For closed shell atoms in interaction the surface $P(q')$ will be more nearly spherical because all possible states of angular momentum are represented in the outer shell. We do not have any molecular binding. Thus it seems reasonable that the statistical field theory should best represent what happens when closed shell atoms come into interaction.

If we made a more realistic *a priori* assumption of form regarding $P(q')$ we could obtain better results from the statistical field. But seemingly this would involve constructing wave functions and evaluating the density matrix so that it is probably best to go directly to some better treatment such as is obtained from the Hartree-Fock equations.

6. ACKNOWLEDGMENTS

This investigation was proposed as a thesis topic by L. H. Thomas, who gave valuable assistance when necessary. The computing was made possible through the cooperation of the Pure and Applied Science Departments of International Business Machines Company. The coefficients required to represent the statistical field equations in curvilinear coordinates (*cf.* Sec. 3 and reference 13) were computed on the IBM Selective Sequence Electronic Calculator ("SSEC"). The coding for this was done by J. Backus and R. Cornish. The statistical field equations themselves were solved on the IBM Electronic Data Processing Machines (type 701) located at the IBM Poughkeepsie Laboratories. In preparing the work for the "701" the author received much valuable assistance from D. W. Ladd and D. Macmillan.

Effects of the Bending Mode of Vibration on the Hyperfine Structure of ICN†

ALI JAVAN
Physics Department, Columbia University, New York, New York
(Received April 20, 1955)

The bending mode of vibration in ICN introduces an asymmetry in the electric field gradient at the iodine nucleus. From an observed variation in the l-type doublet splittings of various hyperfine components for the $v_2=1$ state, the asymmetry parameter is determined as $\eta=0.0087$. It is found that the measured splittings can be accounted for fully by including second-order perturbation terms arising from field asymmetry. The ratio $(eqQ)_{v=0}/(eqQ)_{v_2=1}$ is measured as 1.0039. This ratio and the observed value of η are discussed in terms of the molecular bending and I−C bonding electrons.

INTRODUCTION

PURE rotational transitions in the linear molecule ICN show a rather large hyperfine structure due to the electric quadrupole interaction of the iodine nucleus. In the first excited state of the bending mode of vibration, $v_2=1$, each component of the hyperfine pattern appears as a doublet. This doubling is caused by a removal of degeneracy in the $l=\pm1$ state where l is the quantum number of the internal angular momentum due to vibration. Removal of degeneracy is to a large extent caused by the well-known l-type doubling effect which originates from a rotation-vibration interaction. In a given rotational transition, the contribution from this source to the doublet splitting, being independent of the hyperfine interaction, is equal for all components of the pattern.

A further contribution to the doublet splitting may arise from the quadrupole hyperfine interaction as a result of an asymmetry in the average of the gradient of the electric field at the nucleus. Such an effect causes a variation of the doublet splitting in various components of the hyperfine structure. A preliminary report on the observation of the effect has already appeared in the form of a short abstract.[1] In this paper a more detailed discussion is presented with some further interpretation of the observed hyperfine structure.

FIELD ASYMMETRY

The effect of the asymmetry in the electric field gradient in a linear molecule can be treated in a way similar to that of a slightly asymmetric rotor.[2,3] The asymmetry arises from the bending mode of vibration. The matrix element of the quadrupole hyperfine interaction energy between the $l=+1$ and $l=-1$ states is

$$\langle J, l=\pm1, F|H_Q|J, l=\mp1, F\rangle=-\tfrac{1}{2}\eta eqQf(J,F),$$

where $f(J,F)$ is the Casimir function as tabulated by Townes and Bardeen[4] and where

$$\eta=\frac{\langle\partial^2V/\partial y^2\rangle-\langle\partial^2V/\partial x^2\rangle}{q},\quad q=\left\langle\frac{\partial^2V}{\partial z^2}\right\rangle.$$

The (x,y,z) system of axes is a right-handed system with the z direction along the axis of the molecule and the y direction in the plane of the bent molecule. It should be noted that the z direction is uniquely determined with respect to the bent molecule by the linear momentum and Eckart's condition.[5] In terms of an asymmetric molecule, the x, y, and z axes can be thought to represent, within a good approximation, the three principal axes of inertia with the x axis corresponding to the largest moment of inertia and the z axis the smallest. The averages in the expression for η have to be taken over electronic as well as vibrational motions.

Inclusion of the above term in the doublet splitting, $\Delta\nu$, of the rotational transition $J+1\leftarrow J$ gives, for $F_f\leftarrow F_i$

$$\Delta\nu=2q_l(J+1)+\eta eqQ[f(J+1,F_f)-f(J,F_i)],\quad(1)$$

where the first term is due to the l-type doubling effect and q_l is the l-doubling constant.

Earlier observation[6] on the hyperfine structure of ICN for the state $v_2=1$ failed to detect the presence of a field asymmetry, η, at the nucleus of iodine. The reason was that in the few hyperfine components of the $J=4\leftarrow3$, $v_2=1$, studied by these investigators, the factors multiplying ηeqQ in expression (1) for $\Delta\nu$ were particularly small so that the contributions from this source were masked by their experimental errors. An examination of the hyperfine structure for the same rotational transition reveals that there are some components in the pattern whose doublet splittings show rather large variations A few of these components were chosen for study and determination of the constants of the spectrum.

The presence of an electric quadrupole interaction due to the nitrogen nucleus introduces coupling of the nitrogen spin to the rotation of the molecule. However,

† Work supported jointly by the Signal Corps, the Office of Naval Research, and the Air Research and Development Command.

[1] A. Javan and C. H. Townes, Phys. Rev. **86**, 608 (1952).

[2] C. H. Townes and A. L. Schawlow, *Microwave Spectroscopy* (McGraw-Hill Book Company, Inc., New York).

[3] G. Knight and B. T. Feld, Massachusetts Institute of Technology Research Laboratory of Electronics, Report No. 123, 1949 (unpublished).

[4] J. Bardeen and C. H. Townes, Phys. Rev. **73**, 97 (1948).

[5] C. Eckart, Phys. Rev. **47**, 552 (1935); see also reference 2.

[6] Townes, Holden, and Merritt, Phys. Rev. **74**, 1113 (1948).

since the quadrupole coupling constant of this nucleus is about 800 times smaller than that of the iodine, it causes only slight splittings in the hyperfine pattern. In cases where the nitrogen quadrupole coupling could be resolved, it was found that the nitrogen structures in both members of a given doublet were identical. This indicates that the effect of an asymmetry in the electric field gradient at the nitrogen nucleus is smaller than the experimental errors. Thus the quantity $\Delta\nu$ is independent of the coupling of the nitrogen nucleus and can be designated by the final and initial values of F_1, where F_1 is the quantum number corresponding to the resultant angular momentum of J and the iodine spin Table I contains a list of the measured $\Delta\nu$ and corresponding theoretical values given by (1) The constants q_I and ηeqQ are $q_I=2.643$ Mc/sec; $\eta eqQ=21.00$ Mc/sec. Small discrepencies in the $F_1=7/2\leftarrow 5/2$ and $F_1=5/2\leftarrow 3/2$ transitions, noted in the last column of Table I, are due to a second-order effect involving the field asymmetry which will be discussed shortly.

Resolution of the nitrogen quadrupole splitting makes it possible to measure the coupling constant eqQ of iodine to good accuracy and to determine its variation with the vibrational quantum number v_2. Table II contains a list of the absolute frequencies of the measured hyperfine components in the ground vibrational state and the higher frequency components of the l-doublets, l_2, in the first excited state of the bending mode. Since the nitrogen coupling constant is small, the quantum number F_1 remains approximately a good quantum number. The designation of the states can be, made by the values of F_1 and F where F is the quantum number of the total angular momentum resulting from addition of the angular momentum corresponding to F_1 and the nitrogen spin. However, in the rotational state $J=3$ there is an accidental degeneracy between the levels $F_1=9/2$ and $F_1=7/2$. The perturbation due to the nitrogen coupling causes a considerable admixture of these two states. This fact prevents use of the F_1 quantum number to designate the state of the molecule for these two levels. In Table II these states are denoted by the total angular momentum F in parentheses followed by subscripts 1 and 2. $(F)_1$ is taken to have lower energy than $(F)_2$. The theoretical frequencies listed in Table II are calculated by using a first and second order effects due to the quadrupole coupling of the iodine nucleus. The following coupling constants are determined: For iodine $eqQ=-2420.3\pm0.2$ Mc/sec in the $v=0$ state, $eqQ=-2410.85\pm0.2$ Mc/sec in the $v_2=1$ state giving a ratio $(eqQ)_{v=0}/(eqQ)_{v_2=1}=1.0039\pm0.0001$. For nitrogen $eqQ=-3.8$ Mc/sec.

Using the value of eqQ obtained here for iodine in the $v_2=1$ state, the asymmetry parameter is found to be $\eta=0.0087$. The lines reported in reference 6 agree well with the above constants. It should be pointed out that the presence of the terms in η does not effect the center of a doublet. The asymmetry ηeqQ can conveniently be eliminated in obtaining eqQ by treating the center of the doublets in the hyperfine structure as in the case of no field asymmetry.

TABLE I. Splittings of some of the hyperfine components in the $J=4\leftarrow 3$, $v_2=1$ transition.

$F_{1'}\leftarrow F_1$	Measured	Theoretical (first order)	Difference
$11/2\leftarrow 11/2$	24.715±0.020	24.715	0
$7/2\leftarrow 9/2$ $7/2\leftarrow 7/2$	19.840±0.035	19.840	0
$5/2\leftarrow 7/2$	17.511±0.020	17.549	0.038
$3/2\leftarrow 5/2$	17.310±0.020	17.260	−0.050

TABLE II. Observed frequencies of certain hyperfine components of the ground and $v_2=1$ excited state of ICN. Only the higher-frequency components (l_2) of the l-type doublets are listed, since the lower-frequency components (l_1) can be obtained from these and Table I.

$F_{1'}\leftarrow F_1$	$F_{1'}\leftarrow F_1$	Observed	Calculated
$v_2=1$, l_2:			
$11/2\leftarrow 11/2$	$13/2\leftarrow 13/2$ $11/2\leftarrow 11/2$ $9/2\leftarrow 9/2$	25567.571	25567.58
	$5/2\leftarrow (7/2)_1$	26216.380	216.44
	$7/2\leftarrow (9/2)_1$	26216.771	216.79
	$5/2\leftarrow (7/2)_2$	26217.428	217.49
$3/2\leftarrow 5/2$	$3/2\leftarrow 5/2$ $3/2\leftarrow 3/2$	26265.548	26265.59
	$5/2\leftarrow 5/2$ $5/2\leftarrow 7/2$	26265.210	26265.19
	$1/2\leftarrow 3/2$	26265.610	26265.64
$v=0$			
$11/2\leftarrow 11/2$	$13/2\leftarrow 13/2$ $9/2\leftarrow 9/2$	25393.517	25393.48
	$11/2\leftarrow 11/2$	25393.776	25393.75
$5/2\leftarrow 7/1$		26217.022	
$3/2\leftarrow 5/2$	$3/2\leftarrow 5/2$	26248.971	26249.05

EFFECTS OF SECOND-ORDER IN THE FIELD ASYMMETRY

As is pointed out in reference 6, the second-order corrections for the quadrupole coupling of the iodine nucleus is quite appreciable in the ICN molecule. In the absence of a field asymmetry the correction is of the order $(eqQ)^2/B\times10^{-3}$ Mc/sec. This term arises from matrix elements of the quadrupole interaction between states $\langle J;l;F_1|$ and $\langle J\pm1,\pm2;l;F_1|$. The inclusion of a field asymmetry gives rise to new matrix elements between states $\langle J;l;F_1|$ and $\langle J\pm1,\pm2;l\pm2;F_1|$. Since these elements are off-diagonal in the quantum number l they effect only the doublet splittings. Therefore, the centers of the doublets are independent of η even to a second-order calculation. The second-order contribution to the splitting from the field asymmetry alone is of the order $[\eta^2(eqQ)^2/B]\times10^{-3}$ Mc/sec and, because of the smallness of η, is negligible. However, the presence of a

second-order correction due to the cross terms in elements diagonal and off diagonal in l introduces a correction of the order $[\eta(eqQ)^2/B]\times 10^{-3}$ which is large enough to account for a small deviation of the observed $\Delta\nu$ from the first-order expression mentioned earlier.

The part of the matrix element $\langle J; l; F_1|H_Q|J\pm 1, \pm 2; l\pm 2; F_1\rangle$ which depends essentially on the overall rotation of the molecule can conveniently be factored out.[7] The factor depending on the rotation is

$$\langle J; l; M_J=J|\nabla E_{zz}|J\pm 1, \pm 2, l\pm 2; M_J=J\rangle.$$

The similarity of the treatment of the quadrupole interaction with that of an asymmetrical top rotor was mentioned earlier. In the latter case the quantum number K takes the place of l, and we shall use K instead of l below. The formula given here can be applied to a second-order calculation in the case of an asymmetric rotor or for a symmetric top molecule with the interacting nucleus located off the symmetry axis.

The above matrix element can be evaluated either by a direct integration over the rotational wave function or the use of Racah's coefficients. The following expressions are obtained:

$$\langle J, K, J|\nabla E_{zz}|J+1, K-2, J\rangle = \frac{-\eta eqQ}{2(J+2)(J+1)} \times\left[\frac{(J-K+1)(J-K+2)(J-K+3)(J+K)}{2J+3}\right]^{\frac{1}{2}},$$

$$\langle J, K, J|\nabla E_{zz}|J+1, K+2, J\rangle = \frac{+\eta eqQ[J+K+3]}{2(J+2)(J+1)} \times\left[\frac{(J+K+2)(J+K+1)(J-K)}{(2J+3)(J+K+3)}\right]^{\frac{1}{2}},$$

$$\langle J, K, J|\nabla E_{zz}|J+2, K-2, J\rangle = \frac{-\eta eqQ}{2(J+2)(2J+3)} \times\left[\frac{(J-K+1)(J-K+2)(J-K+3)(J-K+4)}{(J+1)(2J+5)}\right]^{\frac{1}{2}},$$

$$\langle J, K, J|\nabla E_{zz}|J+2, K+2, J\rangle = \frac{-\eta eqQ[(J+K)^2+7(J+K)+12]}{2(2J+3)(J+2)} \times\left[\frac{(J+K+1)(J+K+2)}{(J+1)(2J+5)(J+K+3)(J+K+4)}\right]^{\frac{1}{2}}.$$

[7] J. K. Bragg, Phys. Rev. 74, 533 (1948).

The factors multiplying these expressions to give the total matrix elements can be found in reference 7. The correction to the splitting of the $(J, F_1, K=\pm 1)$ energy level is

$$\delta E=2\sum_{J,K'}\frac{\langle J, K=+1, F_1|H_Q|J', K', F_1\rangle \times\langle J', K', F_1|H_Q|J, K=-1, F_1\rangle}{E_J-E_{J'}}.$$

An evaluation of this quantity for the doublets listed in Table I gives a correction of -58 kc/sec for the transition $3/2\leftarrow 5/2$, 36 kc/sec for $5/2\leftarrow 7/2$, and only about 10 kc/sec for the rest of the transitions.

Thus, the discrepancies listed in Table I is fully accounted for by this second-order effect and we can conclude that the inclusion of the asymmetry parameter η at the nucleus of iodine alone is sufficient to account for the observed variations of the doublet splittings in the hyperfine structure of ICN.

CALCULATION OF η AND VIBRATIONAL DEPENDENCE OF eqQ

Let us first ignore the effect of the molecular bending on the electronic structure of the molecule. When the molecule is bent, the axis of symmetry for the cylindrical charge distribution around the iodine nucleus no longer coincides with the principal axis z of inertia of the molecule. This introduces an asymmetry into the gradient of the electric field in the coordinate system corresponding to the principal axes of inertia which depends on the degree to which the molecule is bent. Hence the asymmetry must be averaged over the vibrational state to give the quantity η of Eq. (1). The tensor for the electric field gradient is diagonal when expressed in a system of axes (x',y',z') with z' in the direction of the axis of cylindrical charge distribution. In this system of axes, the tensor depends on only one parameter because of the relations:

$$\left\langle\frac{\partial^2 V}{\partial x'^2}\right\rangle=\left\langle\frac{\partial^2 V}{\partial y'^2}\right\rangle=-\frac{1}{2}\left\langle\frac{\partial^2 V}{\partial z'^2}\right\rangle.$$

A transformation of this tensor to the principal axes of inertia of the molecule, (x,y,z), gives the following relations:

$$\left\langle\frac{\partial^2 V}{\partial y^2}\right\rangle-\left\langle\frac{\partial^2 V}{\partial x^2}\right\rangle=\tfrac{3}{2}q'\langle\sin^2\theta\rangle,$$
$$\left\langle\frac{\partial^2 V}{\partial z^2}\right\rangle=q'(1-\tfrac{3}{2}\langle\sin^2\theta\rangle), \qquad (2)$$

where θ is the angle between z and z' and where q' is the average of $\partial^2 V/\partial z'^2$ over electronic wave function.

From the observed value of $\eta=0.0087$ and the above equations, one obtains $\langle\sin^2\theta\rangle=0.0058$.

Under the assumption that the axis of symmetry for

the charge distribution is in coincidence with I−C direction, one may calculate η directly as follows:

The angle θ between the I−C direction and the z axis can be expressed in terms of the distance r of the carbon nucleus from the line through I and N. Using the known interatomic distances of ICN and the conservation of the linear momentum, and Eckart's condition one obtains for small values of θ, $\theta=0.385r$. In this relation, θ is in radians if r is expressed in angstrom. Since the angle θ remains small in the course of vibration, the value of the angle can be substituted for $\sin\theta$. Thus, for the evaluation of $\langle\sin^2\theta\rangle$ it is necessary to estimate the average of r^2.

The vibrational wave function in the vth excited state of the bending mode of oscillation with internal angular momentum l is

$$\psi_{v,l}=Ne^{il\chi}\exp(-\rho^2/2)\rho^{|l|}L_{\frac{1}{2}(v+|l|)}^{|l|}(\rho^2),$$

where $L_{\frac{1}{2}(v+|l|)}^{|l|}$ is the associated Laguerre polynomial and where χ and ρ are defined by the normalized doubly degenerate normal coordinates of oscillation, Q_1 and Q_2, as:

$$Q_1=\rho\cos\chi,\quad Q_2=\rho\sin\chi.$$

The quantity ρ in terms of the distance r takes the form αr with the proportionality constant α given by $2\pi(\mu\nu/h)^{\frac{1}{2}}$ where ν is the normal frequency of the bending mode of vibration and where the quantity μ, the reduced mass, is given by

$$\mu=\frac{m_2(m_1+m_2+m_3)I^e}{m_1m_3(l_{1,2})^2\left(\left[\frac{m_2}{m_1}(l_{2,3})^2/(l_{1,3})^2\right]+\left[\frac{m_2}{m_3}(l_{1,2})^2/(l_{1,3})^2\right]+1\right)}.$$

In this equation, I^e is the equilibrium moment of inertia of the molecule; the indices 1, 2, and 3 refer to the three nuclei with subscript 2 standing for the middle nucleus; the m_i's are the nuclear masses and $l_{i,j}$'s the internuclear distances. For ICN, one obtains $\mu=8.858$. The angle χ in the above wave function is the inclination of the plane of the bent molecule with respect to a fixed plane passing through the Z axis.

Evaluation of the average of r^2 over this wave function gives

$$\langle v_2=1,\ |l|=1|r^2|v_2=1,\ |l|=1\rangle=2/\alpha^2,$$

yielding $\eta=0.0051$.

This calculated value is only 60% of the value 0.0087 obtained from the experimental results. Such a discrepancy is considerably outside of experimental error and indicates clearly that the assumption of the I−C direction taken as the axis for a cylindrical charge distribution is not justified.

For an explanation of this discrepancy, we propose the following two effects:

1. An increase in the value of η over the above calculated value can be expected if one introduces the notion of a "bent bond" in the iodine-carbon bonding electrons. For such a bond, the axis of the charge distribution around the iodine nucleus deviates from the I−C direction, hence resulting in such increase.

2. The electron bond configuration in the ICN molecule is known to involve the two structures I−C≡N and $I^+=C=N^-$, with about 13% importance for the second resonating structure. If one ignores the effect of bending on the electronic configuration, a cylindrically symmetric charge distribution for the iodine-carbon bonding electron is required and therefore the π orbitals in the second resonating structure necessarily will consist of 50% π_x and 50% π_y orbitals. The presence of the bending mode of vibration, however, introduces a nonequivalence in the orbitals in the plane of bending and perpendicular to this plane, hence π_x and π_y orbitals may take an unequal population thus destroying the cylindrical symmetry of the charge distribution. This effect also can result in an increase of the value of η.

Unfortunately, our experimental data cannot determine separately the extent to which each of these effects occur. In order to show the meanings of our measured quantities in terms of the above considerations, first let us treat the case where the electron distribution around the iodine nucleus remains cylindrically symmetric, but a bent bond is introduced in the I−C bond as a result of molecular bending:

The quantity $\langle\sin^2\theta\rangle=0.0058$ evaluated above from the measured value of η corresponds to the angle θ_1 between the axis of symmetry for the electric charge distribution and the axis of the molecule. The corresponding average for the angle θ_2 formed by the I−C direction can be calculated from $\langle r^2\rangle$ as 0.0034. The difference between these two values indicates that under the assumption of a cylindrically symmetric charge distribution a bent bond necessarily occurs as a result of the bending mode of vibration. To a good approximation θ_1 can be taken as a linear function of the normal coordinates of the bending mode. This results in a proportionality between θ_1 and θ_2 with the proportionality constant $(\langle\theta_1^2\rangle/\langle\theta_2^2\rangle)^{\frac{1}{2}}=1.30$. The value of $q=\langle\partial^2V/\partial z^2\rangle$ according to the Eq. (2) depends on $\langle\sin^2\theta\rangle$. This dependence accounts for the observed variation in the coupling constant eqQ with the vibrational quantum number v_2. From (2) one obtains:

$$((eqQ)_{v=0}/(eqQ)_{v_2=1})=\frac{q'_{v=0}}{q'_{v_2=1}}\frac{1-\frac{3}{2}\langle\sin^2\theta_1\rangle_{v=0}}{1-\frac{3}{2}\langle\sin^2\theta_1\rangle_{v_2=1}}. \quad (3)$$

$\langle\sin^2\theta_1\rangle$ can be estimated by observing the fact that θ_1 is linearly related to θ_2 and the average of θ_2^2 over the zero point vibration is half as large as that for the $v_2=1$ state. Therefore

$$\langle\theta_1^2\rangle_{v=0}=\tfrac{1}{2}\langle\theta_1^2\rangle_{v_2=1},$$

which gives

$$(1-\tfrac{3}{2}\langle\sin^2\theta_1\rangle_{v=0})/(1-\tfrac{3}{2}\langle\sin^2\theta_1\rangle_{v_2=1})=1.0040.$$

From this ratio and the observed ratio

$$[(eqQ)_{v=0}/(eqQ)_{v_2=1}]=1.0039\pm0.0001,$$

one obtains for $q'_{v=0}/q'_{v_2=1}$ a value of very nearly unity.

Let us consider next the 2nd effect where an unequal population in the π_x and π_y orbitals in the absence of a bent bond causes the increase in η:

In this case due to the fact that $\langle\partial^2V/\partial x'^2\rangle \neq\langle\partial^2V/\partial y'^2\rangle$, Eqs. (2) are modified and take the form:

$$\langle\partial^2V/\partial y'^2\rangle-\langle\partial^2V/\partial z'^2\rangle=\tfrac{1}{2}q'[3\langle\sin^2\theta\rangle+\eta'(2-\langle\sin^2\theta\rangle)],$$

$$\langle\partial^2V/\partial z'^2\rangle=\tfrac{1}{2}q'[2-3\langle\sin^2\theta\rangle+\eta'\langle\sin^2\theta\rangle],$$

where

$$\eta'=\left(\left\langle\frac{\partial^2V}{\partial y'^2}\right\rangle-\left\langle\frac{\partial^2V}{\partial x'^2}\right\rangle\right)\Big/\left\langle\frac{\partial^2V}{\partial z'^2}\right\rangle,$$

and where θ is the angle between the z' direction and the axis z of the molecule with z' in the I−C direction and y' in the plane of bending. Ignoring the small quantity $\eta'\langle\sin^2\theta\rangle$, we obtain $\eta=(3/2)\langle\sin^2\theta\rangle+\eta'$.

The calculated value of $\langle\sin^2\theta\rangle$ and our measured η yields $\eta'=0.0035$.

In terms of the molecular orbitals η' can be interpreted as follows:

The σ bond is cylindrically symmetric and does not contribute to η'. In the resonant structure $I^+=C=N^-$, the wave function for the π electron in terms of p_x and p_y orbitals has the form $\psi=\chi^{\frac{1}{2}}p_x+(1-\chi)^{\frac{1}{2}}p_y$. It can be shown that such wave function gives $\eta'=3(1-2\chi)$. Since this structure has an importance of about 13%, the expected value of η' is $0.39(1-2\chi)$ giving $\chi=0.49$. This means that a 2% decrease in the p_x orbitals for the perpendicular bond at the iodine nucleus can account for the observed value of η.

For the ratio of the quadrupole coupling constant in the ground vibrational state and the first excited state of the bending mode, the Eq. (3) still holds where θ_1 should be taken as the angle formed by the I−C direction and the axis z of the molecule. The average of the square of this angle in the ground and the $v_2=1$ state discussed above yields for $q'_{v=0}/q'_{v_2=1}$ a value again very nearly unity.

The fact that a bent bond can occur together with a nonequivalent p_x and p_y orbitals in the resonant structure $I^+=C=N^-$ implies that the amount of bending and the change in the population of the p_x and p_y orbitals calculated above are the upper limits for these effects. However, in the calculation of the change in the electric field gradient along the axis with the quantum number v_2, either of the above two assumptions yields an undetectable change. The observed variation of the coupling constant eqQ with the quantum number v_2 in ICN, therefore, is mostly a kinematical effect due to the dependence of this quantity on $\langle\sin^2\theta\rangle$.

One should bear in mind that a dependence of eqQ on $\langle\sin^2\theta\rangle$ is not peculiar to linear molecules. For instance, the perpendicular modes of oscillation in the symmetrical top molecules, such as in XYZ_3 molecules, introduces a similar dependence of the coupling constant. In general, a larger variation in eqQ may be expected in these types of oscillation compared to the parallel modes of vibration for which the system of principal axes of the electric field gradient tensor undergoes no rotation with respect to the principal axes of inertia of the molecule.

ASYMMETRY PARAMETER η, AND VIBRATIONAL DEPENDENCE OF eqQ IN BrCN

The rotational transition $J=6\leftarrow5$ in the first excited state of the bending mode of vibration in BrCN has been measured by Tetenbaum.[8] A variation in the l-doublet splittings of various hyperfine transitions was observed by him. However, he did not attempt to analyze his data on the basis of an asymmetry in the electric field gradient. From the data in Table III of his paper, one obtains a value of $\eta=0.0095\pm0.0010$ for either of the two isotopic species in Br. Specifically, this value is calculated from the two transitions $9/2\leftarrow9/2$ and $13/2\leftarrow13/2$, which show the largest variations in their splittings. The agreement of formula (1) of the present paper with the splittings of the rest of the hyperfine transitions of BrCN is satisfactory and to within the reported experimental errors.

Considerations similar to those for ICN above show that for BrCN the asymmetry may be produced by a bent bond with the ratio $\theta_1/\theta_2=1.25$ or by a small amount of asymmetry in the double-bond structure $Br^+=C=N^-$ during bending.

The ratio of the quadrupole coupling constants eqQ in the ground and the first excited state of the bending made of oscillation in BrCN is about the same in both isotopic species of Br and is $(eqQ)_{v=0}/(eqQ)_{v_2=1}=1.004\pm0.001$. For a bent bond one obtains, along the lines discussed for ICN, $q'_{v=0}/q'_{v_2=1}=0.999\pm0.001$ where as in the absence of a bent bond one obtains $q'_{v=0}/q'_{v_2=1}=1.001\pm0.001$.

These results are in consistence with the above observation on the ICN molecule.

ACKNOWLEDGMENTS

The author wishes to acknowledge the encouragement and assistance given by Professor Charles H. Townes in the course of this research.

[8] S. J. Tetenbaum, Phys. Rev. 86, 440 (1952).

Absorption Studies of Helium Metastable Atoms and Molecules

A. V. PHELPS
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received May 5, 1955)

The lifetimes of the metastable states of the helium atom and molecule in pure helium have been determined by using a time-sampling technique for the measurement of the time-varying optical absorption caused by the various metastables. The measured loss of helium singlet metastable atoms indicates destruction after diffusion to the wall, upon collisions with single atoms, and by conversion into triplet metastable atoms in collisions with thermal electrons. The triplet metastable atoms are destroyed upon diffusion and by collisions with two neutral atoms which result in the formation of molecules in the $2^2\Sigma_u^+$ state. These molecules have a natural lifetime of at least 0.05 second and are destroyed upon diffusion to the wall and by collisions with slow electrons or other metastables. The product of the diffusion coefficient and the helium atom density at 300°K is 1.5×10^{19} (cm²/sec)(atom/cc) for the atomic metastables and 1.0×10^{19} (cm²/sec)(atom/cc) for the molecular metastable. The cross sections for the destruction of the singlet atomic metastables are 3×10^{-14} cm² and 3×10^{-20} cm² for collisions with thermal (300°K) electrons and neutral atoms, respectively. The three-body combination coefficient for the conversion of triplet atomic metastables into molecular metastables is 2.5×10^{-34} cc² atom^{-2} sec^{-1} at 300°K.

I. INTRODUCTION

THE processes which determine the rates of destruction of helium metastable atoms in helium were originally studied by Ebbinghaus[1] and have been recently re-examined by Biondi[2] and by Phelps and Molnar.[3] In the latter paper, it is shown that the pressure and temperature variation of the lifetime of helium atoms in the triplet metastable state ($1s\,2s\,^3S_1$) is consistent with the proposal that the metastables are destroyed upon diffusion to the wall of the absorption cell and by collisions with two neutral atoms. It was found that the rate of destruction of helium atoms in the singlet metastable state ($1s\,2s\,^1S_0$) is strongly dependent upon the intensity of the pulsed discharge used to create the metastables. Recently, we have applied time sampling techniques to the measurement of the time-varying optical absorption[4] caused by the metastables and have obtained an increase in sensitivity of roughly two orders of magnitude. The purpose of the experiments described in this paper was to determine the collision processes responsible for the destruction of and conversion between the various metastable states from studies of the time variation of density of the metastables using the improved techniques. Except where otherwise indicated, the destruction of metastables as a result of collisions between pairs of metastables[2] will be assumed to be negligible. Quantitative studies of this process have proved to be quite difficult, but it is hoped that the results can be presented in a latter paper.

II. EXPERIMENT

The ac time sampling technique for the measurement of time-varying optical absorption is described in detail in reference 4. We will give a brief description of the principles of this technique using the block diagram of the apparatus shown in Fig. 1 and the wave forms shown in Fig. 2. The metastable atoms are created by periodically applying a short high-voltage pulse to the electrodes of the absorption cell so as to produce a discharge, Fig. 1(a). The metastables are detected by measuring the absorption of radiation emitted in transitions which terminate on the metastable state of the atom or molecule. The radiation is obtained from a helium source[5] and is collimated and passed through the absorption cell. The desired radiation is separated from the remainder of the helium spectrum by means of an interference filter and focused on the cathode of a photomultiplier. Corresponding to the changes in the metastable density, Fig. 2(b), there are changes in the number of photoelectrons produced at the cathode of the photomultiplier as shown by the dotted curve of Fig. 2(c). If the photoelectric current were amplified and presented on an oscilloscope, the technique described so far would be essentially the same as that used in recent studies.[3] The minimum measurable absorption is limited by the statistical fluctuations or "shot noise" in the number of photoelectrons leaving the photomultiplier cathode. While this noise can be reduced by decreasing the bandwidth of the oscilloscope amplifier, one is limited to a value determined by the bandwidth required to transmit the changing absorption signal. One solution to this problem is to operate the optical system for only a small portion of the afterglow and to average the resultant absorption signal over many cycles of the discharge. Such a time sampling system makes possible a large reduction in the band width of the detector and a consequent reduction in the shot

[1] E. Ebbinghaus, Ann. Physik **7**, 267 (1930).
[2] M. A. Biondi, Phys. Rev. **82**, 543 (1951) and **88**, 660 (1952).
[3] A. V. Phelps and J. P. Molnar, Phys. Rev. **89**, 1203 (1953).
[4] A. V. Phelps and J. L. Pack, Rev. Sci. Instr. **26**, 45 (1955).

[5] The radiation source used in these experiments was either a hot-cathode capillary discharge tube (reference 3) filled with helium to a pressure of 8 mm Hg or the "abnormal" cathode glow of a cold-cathode discharge tube filled to 20 mm. The capillary source emitted only atomic lines while the cold-cathode source emitted approximately equal intensities of atomic and molecular radiation.

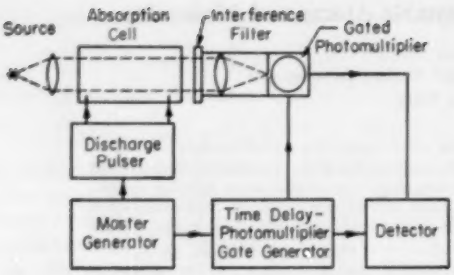


FIG. 1. Schematic of ac time-sampling apparatus for the measurement of time-varying optical absorption. The discharge pulser operates on alternate pulses from the master generator. The pulses from the master generator are delayed by a variable time-delay unit and then supplied to the photomultiplier gate generator and the synchronous rectifier of the detector unit.

noise without sacrificing the ability of the system to resolve changes in the absorption signal.[6] A simple system for "gating" the optical system is to use a continuous source and to apply a short high-voltage pulse to the dynodes of the photomultiplier.

We are now faced with the problem of measuring the small changes in photomultiplier current which are theoretically observable with the time-sampling techniques. One possibility is to "balance out" the signal corresponding to zero absorption either with the output of a second photomultiplier or with a dc signal.[6] A much more convenient technique is to gate the photomultiplier at twice the discharge frequency as shown by the solid curve of Fig. 2(c). Under these conditions, the current pulse immediately following the discharge is reduced as a result of the absorption by an amount large compared to that of the pulse occurring late in the afterglow. The fundamental component of this current waveform, Fig. 2(d), is directly proportional to the difference between the pulse heights and is separated from the undesired signal by means of a narrow-band amplifier and synchronous detector operating at a frequency equal to the repetition rate of the discharge. The variation of the amplitude of this ac signal with the time, t, between the discharge pulse and the first photomultiplier gate gives directly the time variation of the absorption caused by the metastables. The absolute fractional absorption can be obtained by making measurements under conditions in which the absorption is sufficient to cause observable changes in the average photomultiplier current.

The ac time-sampling technique described above provides a hundred-fold increase in sensitivity over methods which do not make use of time-sampling. This increased sensitivity makes it possible to make measurements at average metastable densities which are at least an order of magnitude lower than those previously studied.[8] Under these conditions, the measured absorption can be shown to be directly proportional to the metastable density so that the previous uncertainties as to the correction for nonlinear absorption are eliminated.[7] Measurements at lower metastable densities reduce the destruction of metastables due to collisions with electrons and other metastables. In addition, the increased sensitivity makes possible measurements of the absorption by metastables having low effective absorption coefficients, e.g., the helium metastable molecule.

The electron density present during the decay of the helium singlet metastables was calculated from measurements of the shift in the resonant frequency of a microwave cavity surrounding the absorption cell.[8] The arrangement of the microwave apparatus was the same as that described by Biondi[9] except that the discharge was initiated with a high-voltage pulse rather than with a pulse of microwave power from a magnetron. The vacuum system used in these experiments is typical of those used in this laboratory.[10] The ultimate pressure was 10^{-9} mm Hg or less, and the rate of rise of pressure in the isolated system was 10^{-9} mm Hg/min or less. The gas samples used to obtain the lifetimes of the atomic metastables were "reagent grade" helium obtained from Air Reduction Sales Company. Studies of the effect of known concentrations of neon on the lifetime of the molecular metastable showed that these samples contained neon concentrations of the order of two parts in 10^6. The neon was not removed by the pulsed discharge, as were small concentrations of other impurities, such as argon. The final studies of the molecular metastables were made with helium samples ob-

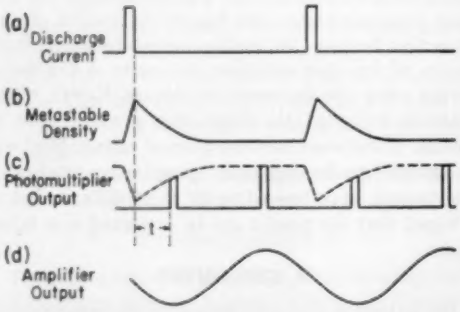


FIG. 2. Wave forms for ac time-sampling technique for the measurement of time-varying optical absorption.

[6] The original experimental arrangements of Meissner and others were of this class by virtue of the use of a pulsed source of radiation. See reference 7. Apparently they were unable to reach the limit of sensitivity set by shot noise because of the difficulties in the measurement of small changes in intensity with photographic plates and the fluctuations in the source intensity under pulsed operation. D. J. Eckl, Can. J. Phys. **31**, 804 (1953) has overcome these particular difficulties by using two photomultipliers as detectors. One photomultiplier is used to "balance out" the signal due to the unabsorbed light.

[7] A. C. G. Mitchell and M. K. Zemansky, *Resonance Radiation and Excited Atoms* (University Press, Cambridge, 1934), Chap. IV.
[8] M. A. Biondi and S. C. Brown, Phys. Rev. **75**, 1700 (1949).
[9] M. A. Biondi, Rev. Sci. Instr. **22**, 500 (1951).
[10] D. Alpert, J. Appl. Phys. **24**, 860 (1953).

tained by a careful evaporation of liquid helium. These samples contained less than two parts in 10^7 of neon.

All of the data presented in this paper were obtained at room temperature, i.e., 300±5°K. If we make the extreme assumption that all of the discharge energy appears as heat in the volume of the gas, the calculated temperature rise at the end of the discharge pulse is about 5°C for a helium pressure of one mm Hg and proportionately less at higher pressures.[8,11] Experimentally, there was no observable rise in the temperature of the absorption cell wall. Measurements of the metastable absorption as a function of the intensity of the source showed that the destruction of metastables as a result of excitation to radiating states was negligible.

III. EXPERIMENTAL RESULTS AND THEIR INTERPRETATION

The experimental results will be presented in the order of decreasing energy of the metastable state involved. Such a presentation allows a more natural discussion of the conversion of one metastable state into another. Thus, Fig. 3 shows that the decay in the absorption of the 5016 A line by the 2^1S metastables is accompanied by an increase in the absorption of the 3889 A line by 2^3S metastables. This observation suggests that atoms in the 2^1S state are being converted into atoms in the 2^3S state. Studies of the decay of the 2^1S density, to be discussed below, show that under the conditions of this experiment the rate of 2^1S destruction is determined by the density of electrons present during the afterglow. Accordingly, we propose that the metastables are converted according to the reaction

$$\mathrm{He}(2^1S)+e\rightarrow\mathrm{He}(2^3S)+e+0.79\ \mathrm{ev}.$$

This qualitative picture leads us to consider solutions to the differential equations governing the metastable behavior. It is assumed that the decay 2^1S metastable density, S, is determined by their loss due to diffusion, to collisions with neutral atoms, and collisions with thermal electrons. The differential equation expressing this is

$$\partial S/\partial t=D_S\nabla^2S-\nu(N)S-\beta nS, \qquad (1)$$

where D_S is the diffusion coefficient for the 2^1S metastables, $\nu(N)$ is the frequency of destruction of the 2^1S metastables as a result of collisions with neutral atoms, N, and β is the frequency per electron for the conversion of 2^1S metastables into 2^3S metastables by thermal electrons of density, n.

If we assume that the 2^1S density is distributed in the fundamental diffusion mode and that the time and spatial variation of the electron density can be neglected, the solution to Eq. (1) is

$$S=S_0e^{-\nu_Bt}, \qquad (2)$$

where

$$\nu_B=D_S/\Lambda^2+\nu(N)+\beta n \qquad (3)$$

[11] E. H. S. Burhop, Proc. Phys. Soc. (London) **A67**, 276 (1954).

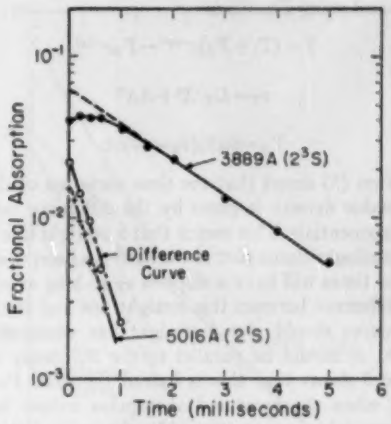


FIG. 3. Time variation of absorption of 5016 A and 3889 A lines by helium singlet (2^1S) and triplet (2^3S) metastables for a pressure of 1.5 mm Hg and $\Lambda^2=0.053$ cm^2. The initial triplet metastable density was approximately 10^{11} metastables/cc.

and Λ is the diffusion length[12] for the absorption cell. The assumption that the electron density is essentially constant during the 2^1S decay is shown to be valid by simultaneous microwave measurements of the time variation of the electron density. The electron density is found to change by less than 20 percent. An analytical solution of Eq. (1) for a cosinusoidal spatial distribution of the electrons can be obtained for the one dimensional case in terms of Mathieu functions. This investigation shows that the cross section calculated from the central electron density is low by about 20 percent. Accordingly, the corrected value of β from the plot of ν_B *vs* n in Fig. 4 will be taken as 3.5×10^{-7} cc/electron-sec and corresponds to an average cross section for 300°K electrons of 3.0×10^{-14} cm^2. Since the collision which converts a 2^1S metastable into a 2^3S metastable is an exchange collision, the principal contribution to the cross section at low electron energies is expected to be due to electrons with zero angular momentum.[13] In this case theory shows that the maximum allowable cross section for conversion by thermal electrons is 4×10^{-14} cm^2.

The differential equation governing the triplet metastable density, T, is

$$\partial T/\partial t=D_T\nabla^2T-AN^2T+\beta nS, \qquad (4)$$

where D_T is the diffusion coefficient for the 2^3S metastables and AN^2 is the frequency for the destruction of the 2^3S metastables in collision with two neutral atoms.[3] The solution to Eq. (4) under the assumption of a spatial distribution given by the fundamental diffusion

[12] The diffusion length, Λ, for a cylindrical absorption cell is given by $1/\Lambda^2=(\pi/L)^2+(2.4/R)^2$, where L and R are the length and radius of the cylinder.

[13] H. S. W. Massey and E. H. S. Burhop, *Electronic and Ionic Impact Phenomena* (Oxford University Press, London, 1952), Chap. III.

mode and using Eq. (2) is

$$T=(T_0+T_S)e^{-\nu_T t}-T_S e^{-\nu_S t}, \tag{5}$$

where

$$\nu_T=D_T/\Lambda^2+AN^2 \tag{6}$$

and

$$T_S=\beta n S_0(\nu_S-\nu_T)^{-1}.$$

Equation (5) shows that the time variation of the 2^3S metastable density is given by the difference between two exponentials. This means that a straight line fitted to a semilogarithmic plot of the 3889 A absorption data at late times will have a slope of ν_T so long as $\nu_S>\nu_T$. The difference between this straight line and the measured curve should give a straight line whose slope is ν_S, i.e., it should be parallel to the 2^1S decay curve. Figure 3 shows that this is indeed the case. Furthermore, when the measured absorption values for the 2^1S metastables are multiplied by the ratio of the theoretical absorption coefficients,[14] as has been done in Fig. 3, the difference density curves agree in magnitude with predictions of Eq. (5) to within 20 percent.

In principle, the pressure variation of the destruction frequency for the 2^1S metastable can be determined by extrapolating the data given in Fig. 4 to zero electron density. However, at the lower electron densities the scatter in the data is usually large enough to obscure small variations in ν_S with pressure. So far, the only usable data on the pressure variation of ν_S was obtained by adding a low-voltage (<1000 volts) pulse lasting for about one millisecond to the breakdown spark in such a way as to terminate the discharge with a very low current and electron density but with a measurable metastable density.[15] Figure 5 shows the pressure variation of ν_S when taken under conditions such that the electron density was less than 10^8 electron/cc. The smooth curve is fitted to the data on the assumption that 2^1S metastables are destroyed by diffusion and by collisions with single neutral helium atoms. Within the accuracy of this data, the diffusion coefficient agrees with that obtained for the 2^3S metastable and is $D_S=440\pm50$ cm²/sec at 1 mm or $D_S N=1.4\pm0.2\times10^{19}$ (cm²/sec)(atom/cc) at 300°K. The average cross section for the volume destruction of the 2^1S metastables is 3×10^{-20} cm². This value agrees with rough estimates of the cross section expected for the destruction of metastables by collision induced radiative transitions.[2,16] Measurements of the absorption by 2^3S metastables under these conditions failed to show the buildup at early times which would be expected if the collision of a 2^1S metastable with a neutral atom resulted in the formation of a 2^3S metastable.

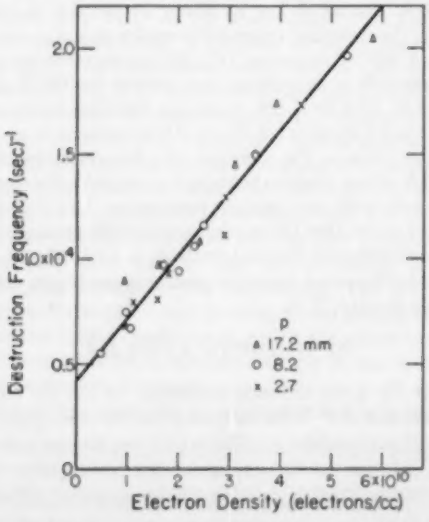


FIG. 4. Measured destruction frequency for helium singlet metastables as a function of the central electron density at various pressures and 300°K. $\Lambda^2=0.053$ cm².

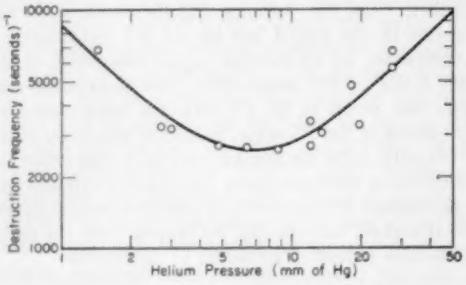


FIG. 5. Measured destruction frequency for helium singlet metastables as a function of helium pressure at 300°K under conditions of negligible electron density. $\Lambda^2=0.053$ cm².

The pressure variation of the destruction frequency for the 2^3S metastable for two absorption cells is shown in Fig. 6. The smooth curves represent fits to the experimental data of Eq. (6) using a single diffusion coefficient. The diffusion coefficient for the 2^3S metastables at 300°K is 470 ± 25 cm²/sec at 1 mm of $D_T N=1.51\pm0.08\times10^{19}$ (cm²/sec)(atom/cc). The value for the three-body combination coefficient is $A=0.26\pm0.03$ sec⁻¹ mm⁻⁴ $=2.5\times10^{-34}$ cc² atom⁻² sec⁻¹ at 300°K. These coefficients are at the upper limit of the experimental error estimated for the previous results.[5] The difference may result from the presence of nonlinear absorption

[14] D. R. Bates and A. Damgaard, Trans. Roy. Soc. (London) A242, 101 (1949).

[15] Calculations of the energy dissipated by the discharge show that the maximum temperature rise in the gas due to this pulse is about equal to that for the breakdown pulse.

[16] The destruction of 2^1S metastables by collision induced radiation was proposed by J. L. Nikerson, Phys. Rev. 47, 707 (1935), in order to explain the presence of a band of continuous radiation near 600 A. The presence of a sharp band edge at a wave number 385 cm⁻¹ larger than that of the 602 A forbidden line was interpreted as indicating the existence of a maximum in the difference in the potential energy curves for the $2^1\Sigma$ and $1^1\Sigma$ states. Such a maximum is predicted from theory in reference 17.

resulting from the relatively large metastable densities required with the less-sensitive techniques used in the earlier work. The small increase in the experimental points above the smooth curve at intermediate pressures is believed to be due to small amounts of neon present in the helium samples used but, in some cases, may indicate a small loss due to collisions of metastables with other metastables or slow electrons. The experimental value for the 2^3S diffusion coefficient is in good agreement with the value of $D_T N = 1.29 \times 10^{19}$ (cm²/sec) (atom/cc) at 300°K calculated from the theoretical diffusion cross section data given by Buckingham and Dalgarno.[17] Although these diffusion measurements are not a very sensitive means for observing the effects of the reflection of metastables from the walls, the fit of the low pressure data to an inverse pressure dependence suggests the reflection coefficient for the metastables at the glass walls is less than 20 percent.[18] The combination

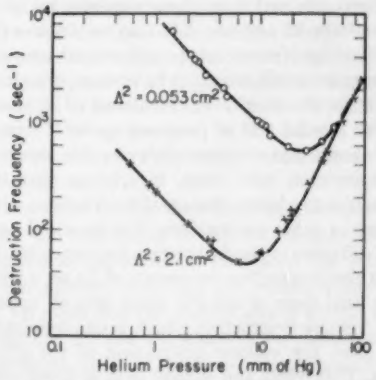


FIG. 6. Measured destruction frequency for helium triplet metastables as a function of helium pressure at 300°K for absorption cells with $A^2 = 0.053$ and $A^2 = 2.1$ cm².

coefficient for the three-body collisions is considerably lower than the corresponding coefficients for similar three-body reactions such as the destruction of argon[3] and mercury[18] metastables. As pointed out previously,[3,11] this low value is consistent with the existence of a long-range repulsive interaction such as that calculated[17] for the metastable states of the helium molecule.

The loss of the 2^3S metastables in three-body collisions with neutral helium atoms suggests that the process results in the formation of an excited triatomic molecule which may radiate before dissociation, dissociate into the neutral atoms, or form a vibrationally stable diatomic molecule by giving up sufficient kinetic energy to one of the neutral atoms. In the latter case, the potential energy curves calculated by Buckingham

[17] R. A. Buckingham and A. Dalgarno, Proc. Roy. Soc. (London) **A213**, 327 and 506 (1952).
[18] A. O. McCoubrey, Phys. Rev. **93**, 1249 (1954).

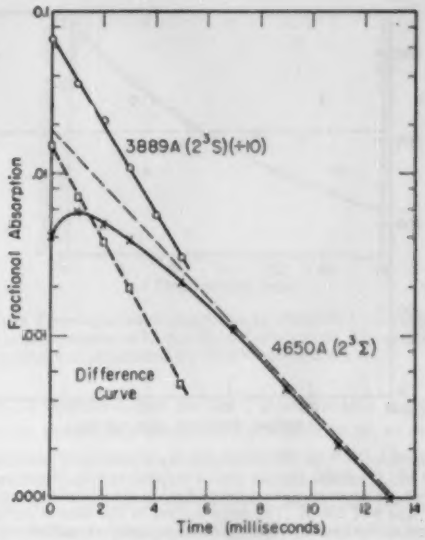


FIG. 7. Time variation of absorption of 3889 A line and 4650 A band by helium triplet atomic (2^3S) and molecular metastables ($2^3\Sigma$) at 300°K for $A^2 = 0.050$ cm² and a helium pressure of 16 mm Hg. The initial 2^3S density was approximately 1.5×10^{11} metastables/cc. The measured values of the 3889 A absorption have been increased by a factor of 2.8 in order to correct for an apparent partial reversal of the 3889 A line by the cold-cathode cource.

and Dalgarno[17] for the $2s\sigma\ ^3\Sigma_u^+$ and $2s\sigma\ ^1\Sigma_u^+$ states indicate that the diatomic molecule should be formed in a high vibrational level of the $2^3\Sigma$ state. If the selection rule forbidding intercombination radiation is sufficiently well obeyed[19,20] the $2^3\Sigma$ molecules will give up their vibrational energy to neutral atoms and reach the lowest vibrational level where they may be detected by measuring the absorption of the 4650 A band of the $3^3\Pi_g \rightarrow 2^3\Sigma_u$ system.[21]

The solid curves of Fig. 7 show the measured fractional absorption for the 3889 A line and the 4650 A band at a pressure of 16 mm in an absorption cell with $A^2 = 0.050$ cm². These curves are interpreted as showing that as the 2^3S metastables are destroyed there is a buildup in the concentration of the $2^3\Sigma$ metastables as a result of the reaction $He(2^3S) + 2He \rightarrow He_2(2^3\Sigma) + He$. We propose that under the conditions of this experiment the $2^3\Sigma$ metastables are destroyed upon diffusion to the

[19] R. Meyerott, Phys. Rev. **90**, 671 (1949). This paper is one of many which give evidence for the existence of a metastable helium molecule. However, the potential energy curve obtained for the ground state of the helium molecule is very difficult to reconcile with the potential energy curve obtained from theory and atomic beam experiments. See, e.g., I. Amdur, J. Chem. Phys. **17**, 844 (1949).
[20] The potential energy curve crossing proposed in reference 11 seems unlikely in view of the failure of the calculated curves to cross (reference 17, p. 338) and the fact that the lowest vibrational level of the $2^1\Sigma$ state is known to be higher than that of the $2^3\Sigma$ state by 0.29 electron volt (reference 21).
[21] G. Herzberg, *Spectra of Diatomic Molecules* (D. Van Nostrand Company, Inc., New York, 1950), p. 536. G. H. Dicke and E. S. Robinson, Phys. Rev. **80**, 1 (1950).

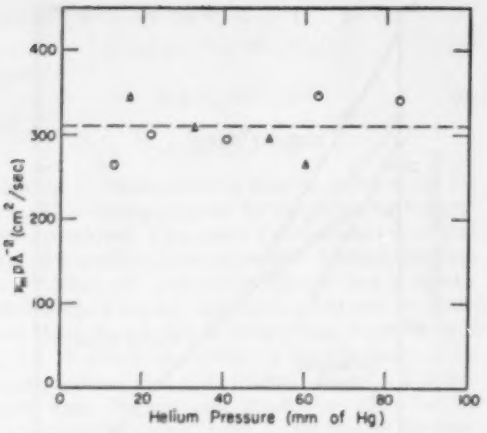


FIG. 8. $\nu_M p\Lambda^{-2}$ for $2^3\Sigma$ metastables as a function of helium pressure for $\Lambda^2=0.050$ cm^2(Δ) and $\Lambda^2=0.015$ cm^2(0). Since $\nu_M p\Lambda^{-2} = D_M p$, its average value (dotted line) is the diffusion coefficient at 1 mm and 300°K. The large scatter in the data is indicative of the difficulties in making measurements at sufficiently low densities to avoid the effects of collisions with slow electrons or other metastables.

wall of the absorption cell. The differential equation governing the time variation of the $2^3\Sigma$ density, M, is

$$\partial M/\partial t = D_M \nabla^2 M + A N^2 T, \tag{7}$$

where D_M is the diffusion coefficient for the $2^3\Sigma$ metastables. Because of the relatively long times and high pressures involved in the studies of the $2^3\Sigma$ buildup we have neglected the time required to dissipate the vibrational excitation energy. One can also neglect the second term in the solution for the 2^3S density given Eq. (5). If we assume a fundamental diffusion mode, the solution to Eq. (7) is

$$M = (M_0 + M_T)e^{-\nu_M t} - M_T e^{-\nu_T t}, \tag{8}$$

where

$$\nu_M = D_M/\Lambda^2 \quad \text{and} \quad M_T = A N^2 (T_0 + T_S)(\nu_T - \nu_M)^{-1}.$$

Equation (8) shows that a straight line fitted to a semilogarithmic plot of the $2^3\Sigma$ absorption data at late times will have a slope equal to ν_M. In addition, the difference between this straight line and the measured curve should be a straight line with the same slope as the 2^3S decay curve, i.e., the difference curve and the 2^3S curve should be parallel. Figure 7 shows that this is indeed the case. The curves can be made to agree in magnitude with theory provided that the ratio of the integrated absorption coefficient for the 3889 A line is 15 times the average absorption coefficient for the 4650 A band.* This ratio contains a factor of 10 to 15 for the ratio of the integrated to average absorption coefficient for the 4650 A band resulting from the distribution of the emission and absorption over a number of transitions involving the various rotational levels of the molecule.[21] In view of our lack of knowledge of the relative values for the integrated absorption coefficients and the uncertainties involved in correcting the measured fractional absorption values for differences in line breadth and fine structure, we conclude that the experimental results are consistent with our hypothesis of complete conversion of the 2^3S metastables to $2^3\Sigma$ metastables in three-body collisions.

Figure 8 shows the results of a series of measurements of the destruction frequency for the $2^3\Sigma$ metastables as a function of pressure in absorption cells with $\Lambda^2=0.050$ cm^2 and 0.015 cm^2. The dashed curve is fitted to the data on the assumption that the metastables are lost by diffusion with a negligible volume loss. The diffusion coefficient obtained from these experiments at 300°K is $D_M=310\pm50$ cm^2/sec at 1 mm or $D_M N=1.0\pm0.2\times10^{19}$ (cm^2/sec)(atom/cc). In order to enhance any volume destruction which might be present, the time variation of the $2^3\Sigma$ density was measured in an absorption cell with $\Lambda^2=2.1$ cm^2 at pressures up to 97 mm. Since Fig. 9 shows that the final decay of the absorption is not exponential with time, it appears that the $2^3\Sigma$ metastables are being destoryed by collisions with slow electrons or other metastables. The expected contribution of diffusion to the destruction frequency is 1.5 sec^{-1}, so that the destruction frequency of 23 sec^{-1} indicated by the final slope of the $2^3\Sigma$ decay sets an upper limit to the volume destruction which is one hundred times lower than the volume destruction of the 2^3S metastables. This result and similar data at lower pressures show that the natural lifetime of the $2^3\Sigma$ metastable state is at least 0.05 sec.

IV. SUMMARY AND DISCUSSION

The results of the measurements of the destruction frequency for the metastable states of helium are summarized in Table I. The agreement between the theoretical and experimental diffusion coefficients for the triplet atomic metastable is as good as can be expected in view of the uncertainties in the energy of interaction for the $2^3\Sigma$ state. Because of the similarities in the in-

TABLE I. Diffusion coefficients and volume loss at 300°K.

Metastable state	Diffusion coefficients at 300°K (DN in (cm^2/sec)(atoms/cc))	Volume loss at 300°K
2^3S	1.4×10^{19}	linear: $(\sigma v)_M=6\times10^{-15}$ cc atom^{-1} sec^{-1}
2^3S	$1.151\pm0.08\times10^{19}$	quadratic: $A=(2.5\pm0.3)\times10^{-31}$ cc^2 atom^{-2} sec^{-1}
$2^3\Sigma$	$1.0\pm0.2\times10^{19}$	natural lifetime>0.05 sec

* *Note added in proof.*—An interferometric study by L. S. Frost shows that the effective temperature of the cold-cathode source is essentially the same as that of the capillary source. This means that the correction factor for the 3889 A absorption in Fig. 7 should have been 1.8 instead of 2.8. The ratio of the integrated absorption coefficient for the 3889 A line to the average absorption coefficient for the 4650 A band is then 10 instead of 15.

teraction curves for the $2^1\Sigma$ and $2^3\Sigma$ states, it is not surprising that their diffusion coefficients are equal to within the experimental error. The indication that the diffusion coefficient for the metastable molecule is smaller than that for the metastable atom is reasonable in view of the larger size of the molecule and the fact the effects of excitation transfer on the diffusion cross section are small.[17]

If the two-body collision destruction of the 2^1S metastables is the result of collision-induced radiation, and if the height of the interaction energy curve is as large as calculated by Buckingham and Dalgarno,[17] then the cross section indicated by our data is surprisingly large. Rough estimates of the probability for collision-induced radiation[2] give cross sections of the order of the observed value only in the absence of a significant repulsive interaction. If the interaction energy for the $2^1\Sigma$ state follows the calculated curve, only one in 10^4 of the thermal collisions could result in an approach to distances of the order of five atomic units where an appreciable radiative transition probability might be expected. This difficulty would not be present if the height of the potential maximum were of the order of that suggested by Nickerson.[16] It should be possible to determine the "activation" energy involved in this process by measuring the variation in the average cross section with the gas temperature.

The low value for the three-body combination coefficient for the conversion of 2^3S metastables into $2^3\Sigma$ metastables as compared to the values for mercury[18] and argon[3] appears to de due to a repulsive interaction for the 2^3S atom and two neutral atoms similar to that found for the $2^3\Sigma$ and $2^1\Sigma$ states of the diatomic molecule. However, there is no reliable theory available for the calculation of three-body combination coefficients. The agreement between the experimental value and the estimates based on the Thompson "recombination" theory[11] and the calculated $2^3\Sigma$ interaction would appear to be fortuitous since the approach of the third helium atom may have a strong effect on the interaction energy of the system. The absence of an appreciable

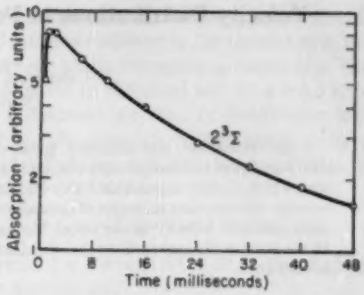


FIG. 9. Time variation of absorption by $2^3\Sigma$ (4650 A) metastables for a helium pressure of 97 mm Hg and $\Lambda^2 = 2.1$ cm^2. The maximum $2^3\Sigma$ density was approximately 2×10^{10} metastables/cc.

three-body destruction for the $2^1\Sigma$ metastable suggests that the three-body interaction is repulsive or, at most, weekly attractive. It is of interest to note that the process of the conversion of an atomic metastable into a diatomic molecular metastable, which we postulate on the basis of direct measurements of optical absorption by the metastables, has been proposed by McCoubrey to explain the production of band fluorescence in mercury[18] and by Colli to explain the time dependence of photoelectric current pulses in a coaxial discharge tube filled with argon.[22]

The measurement of the cross section for the conversion of 2^1S metastables into 2^3S metastables by thermal electrons is probably the first measurement of a "superelastic" cross section for electrons of such low energy. In view of the approach of the experimental value to the possible theoretical maximum it would be very interesting to determine the variation of this cross section with electron energy. This could be done most easily by varying the temperature of the gas.

The author wishes to express his appreciation for many valuable discussions of these problems with his associates in the Atomic Physics Group. He is especially indebted to Mr. J. L. Pack for the design of the electronic components of the time-sampling apparatus.

[22] L. Colli, Phys. Rev. **95**, 892 (1954).

Velocity Distributions in Potassium and Thallium Atomic Beams*†

R. C. MILLER‡ AND P. KUSCH
Columbia University, New York, New York

(Received February 23, 1955)

A high-resolution, high-intensity, spiral velocity selector has been designed for the study of the velocity distributions of the components of atomic and molecular beams. It has been found possible to design oven slits which closely approximate the ideal aperture of kinetic theory. An analysis has been made of the velocity distributions in beams of potassium and thallium over a range of velocity from 0.3 to 2.5 times the most probable velocity in the oven. The agreement between the observed distribution and that deduced, on the basis of the assumptions that the distribution in the oven is Maxwellian and that the aperture is ideal, is very good.

INTRODUCTION

A NUMBER of investigators have measured the velocity distributions of atoms in an atomic beam.[1–8] In a number of cases, in which the results are of at least a moderate precision, significant differences occur, usually on the low velocity side of the spectrum, between the observed distribution and that calculated from the assumption that the distribution in the beam is the consequence of the effusion of atoms through an ideal aperture from an oven in which the velocity distribution is Maxwellian. In view of the limited agreement between observation and theoretical prediction, it is worthwhile to reinvestigate the velocity distribution of atoms in a beam to establish, on experimental grounds, the reality of the Maxwellian distribution of velocities of particles at thermal equilibrium in an isothermal enclosure. The detector in an atomic or molecular beams experiment measures the number of particles incident upon it in unit time. If the slit in the oven is an ideal aperture, then a Maxwellian v^2 volume distribution in the oven becomes a v^3 distribution at the detector. The design of oven slits to produce an ideal aperture and thus give a reproduceable distribution within the beam, derivable from that in the oven, is an important aspect of the present work. The present paper is concerned with the design of a high-resolution velocity selector, with an analysis of the conditions under which the velocity distribution of the beam incident on the detector is derivable from the velocity distribution in the oven and with a velocity analysis of atoms at equilibrium in an isothermal enclosure.

VELOCITY SELECTOR

The velocity selector consists of a solid cylinder on whose surface are cut a large number of helical slots and which is capable of rotation about the cylinder axis. This velocity selector has inherent advantages over a selector which employs two or more toothed wheels operating on the same shaft at a fixed relative angular position and at a variable angular velocity, or operating on two separate shafts at the same fixed angular velocity but at a variable relative angular position.[8] With the toothed wheel selectors, the transmission of a beam by the device is possible when the time of transit of the particles over the length of the selector is equal to the time required for the selector to turn through the interval at whose ends the slot in the first disk and that in the second disk transmit the beam. Evidently, if the wheels contain a large number of closely spaced teeth, atoms within a particular velocity interval may be transmitted for several angular velocities of the rotor; and a particular rotor velocity will allow the transmission of atoms in several velocity intervals. In principle, the occurrence of "side bands" may be eliminated since the slots may be so widely spaced on each disk, relative to the angular interval between the slots on the two disks, that the only significant contribution to the intensity lies within a single velocity range. However, since all velocity selectors reduce the transmitted intensity at a particular central velocity with increasing resolution, the intensity problem may become very serious if additional intensity losses must be incurred when the "side bands" are eliminated. When a single cylinder is used, the spacing between adjacent slots may be as small as mechanical considerations permit. The width of the slots, a determining factor in the resolution, may in turn be as small as is mechanically feasible and consistent with the requirement that the intensity of the transmitted beam over the velocity range to be studied is considerably greater than noise. The present selector is similar to a high transmission, slow neutron velocity

* This work was supported (in part) jointly by the Department of the Army (Signal Corps), the Office of Naval Research, and the Air Research and Development Command.

† Submitted in partial fulfillment (by R. C. M.) of the requirements for the degree of Doctor of Philosophy at Columbia University.

‡ Radio Corporation of America Predoctoral Fellow, 1953–1954. Now at the Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

[1] J. Eldridge, Phys. Rev. **30**, 931 (1927).
[2] B. Lammert, Z. Physik **56**, 244 (1929).
[3] I. F. Zartman, Phys. Rev. **37**, 383 (1931).
[4] C. C. Ko, J. Franklin Inst. **217**, 173 (1934).
[5] V. W. Cohen and A. Ellett, Phys. Rev. **52**, 502 (1937).
[6] Estermann, Simpson, and Stern, Phys. Rev. **71**, 238 (1947).
[7] I. Kofsky and H. Levinstein, Phys. Rev. **74**, 500 (1948).
[8] A. Bennett, Jr., and I. Estermann, Ph.D. thesis, Carnegie Institute of Technology, 1953 (unpublished); A. Bennett, Jr., Phys. Rev. **95**, 608 (A) (1954).

selector.[9] A similar helical velocity selector has also been used in a study of the characteristics of thin metal films as a function of the incident velocity of the deposited atoms.[10]

The notation used in a discussion of the selector is shown in Fig. 1 as are the dimensions§ of the rotor used in the present work. The cylinder was made of Dural because this material allowed the cutting of slots to the necessary depth and at close intervals without significant distortion of adjacent slots. The ends of the slots occur at 0.5° intervals on the end faces. To allow adjustment of the line connecting the source slit and the detector parallel to the cylinder axis, it is necessary to have at least one slot parallel to the axis. Actually, two such slots were cut to avoid dynamical imbalance.

The resolution properties of the velocity selector will be derived under the assumption that both source and detector have infinitesmal dimensions. All molecules which are ultimately detected then travel in a plane containing the cylinder axis, oven slit, and detector. The validity of this assumption will be discussed in more detail later. A molecule of velocity v_0 which goes through the slot without changing its position relative to the sides of the slot will satisfy the equation

$$v_0=\omega L/\phi_0, \tag{1}$$

in which ϕ_0 and L are fixed, and ω is the angular velocity.

Since the slots have a finite width, a range of velocities from $v_{max}=\omega L/\phi_{min}$ to $v_{min}=\omega L/\phi_{max}$ will be transmitted by the rotor for a fixed ω. The limiting velocities correspond to molecules which enter the slot at one wall and leave the slot at the opposite wall, so that

$$v_{max}=v_0(1-\gamma)^{-1} \quad \text{and} \quad v_{min}=v_0(1+\gamma)^{-1}, \tag{2}$$

where $\gamma=l/(r_0\phi_0)$ and r_0 is the mean radius. In the present apparatus, $\gamma=0.05033$, a fixed property. A variation of the radius by a few percent will have a negligible effect on the limiting values of v and thus, the mean value, r_0, may be used. The admittance A, the ratio of the effective slot aperture for the velocity v to that for the velocity v_0 is

$$A=1-|(v_0/v)-1|/\gamma, \tag{3}$$

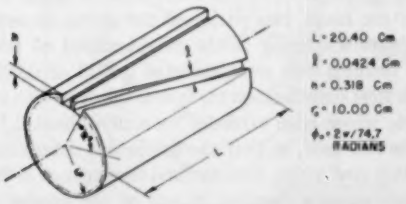


FIG. 1. Schematic diagram of velocity selector.

[9] J. Dash and H. Sommers, Jr., Rev. Sci. Instr. **24**, 91 (1953).
[10] H. Levinstein and H. Crane, Phys. Rev. **69**, 679 (A)(1946).
§ *Note added in proof.*—The dimension $L=25.40$ cm.

where v lies between the limits v_{max} and v_{min}. Over the limited range of v allowed in the present case, the admittance is very nearly triangular in terms of v. The resolution calculated in the usual way at $A=0.5$ is $\Delta v/v_0=\gamma$.

If it is assumed the velocity distribution incident on the rotor, I_0dv, does not change significantly over the range of velocities admitted for a given ω, then the transmitted intensity is

$$I_v=-I_0v_0\gamma^{-1}\ln(1-\gamma^2)\cong I_0v_0\gamma \tag{4}$$

to powers of γ through γ^2. I_0dv is the modified Maxwellian distribution,

$$I_0dv=2\alpha^4v^3\exp(-\alpha^2v^2)dv,$$

where $\alpha^2=m/2kT$, and the distribution has been normalized to unity. Then Eq. (4) becomes

$$I_v=2\gamma\alpha^4v_0^4\exp(-\alpha^2v_0^2). \tag{5}$$

Equation (5) becomes exact as the slot width becomes infinitesimal. It will be shown that this expression is sufficiently close to the much more cumbersome exact expression so that it can be used as the theoretical distribution.

To obtain an exact expression for I_v, the variation of the incident distribution, I_0 over the admittance must be included. Then:

$$\begin{aligned}I_v=&(\gamma^{-1}-1)\exp[-x_0^2/(1-\gamma)^2]\\&+(\gamma^{-1}+1)\exp[-x_0^2/(1+\gamma)^2]\\&-2\gamma^{-1}\exp(-x_0^2)+x_0\gamma^{-1}\\&\times\left\{\int_{x_0}^{x_0/(1-\gamma)}\exp(-x^2)dx\right.\\&\left.+\int_{x_0}^{x_0/(1+\gamma)}\exp(-x^2)dx\right\},\end{aligned} \tag{6}$$

where $\alpha v_0=x_0$. This expression is unwieldy and gives no insight into the general nature of I_v. Table I contains

TABLE I. A comparison of the intensity distribution after analysis by the rotor as calculated in several approximations.

$x_0=\alpha v_0$	Exact intensity Eq. (6)	Expansion value of intensity Eq. (7)	One-term γ^2 intensity Eq. (5)
0.000	0.00	0.00	0.00
0.200	0.62	0.57	0.57
0.400	0.80	0.81	0.81
0.600	3.36	3.36	3.34
0.800	8.01	8.00	7.98
1.000	13.62	13.61	13.59
1.200	18.16	18.15	18.15
1.300	19.45	19.47	19.47
1.414	19.98	19.97	20.00
1.500	19.67	19.68	19.71
1.600	18.69	18.68	18.72
1.800	15.15	15.15	15.19
2.000	10.79	10.81	10.83
2.200	6.84	6.84	6.84
2.400	3.87	3.87	3.86

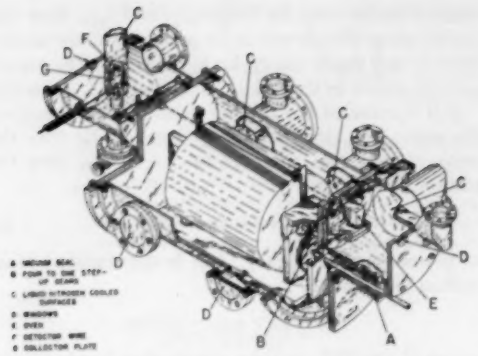


FIG. 2. Schematic diagram of the apparatus designed to measure velocity distributions.

some values of I_v calculated from Eqs. (5) and (6). Both sets of values have been multiplied by a constant factor to make the maximum intensity, $I_{v(\max)}$, of Eq. (5) equal to 20.00. Table I demonstrates that the simple expression, Eq. (5), can be used for calculating theoretical distributions to be compared to the experimental data in the present work, as the average experimental uncertainty is about one percent of $I_{v(\max)}$.

The intensity at the detector, I_v, may be expanded as a series in $(x_0)^{2n}\exp(-x_0^2)$, where n is an integer greater than or equal to two. Then the leading terms are

$$I_v = 2\gamma x_0^4 \exp(-x_0^2)[1+2.5\gamma^2 - 13\gamma^2 x_0^2/6+\gamma^2 x_0^4/3+\cdots] \quad (7)$$

in which the next term in x_0^{10} has a coefficient no larger than γ^3. Table I contains values of I_v calculated from Eq. (7) using terms through x_0^8.

Finite source and detector dimensions produce a beam not all of whose elements are parallel to the axis of the rotor. The effect of the finite vertical dimension of the beam is negligible since for an angle of elevation, θ, of a beam element with respect to the axis, the analysis of $v\cos\theta$ is made and $\cos\theta$ differs only trivially from 1. The center of the admittance curve for non-parallel beam elements which results from the finite horizontal dimensions, or widths, of the oven slit and detector wire is at either higher or lower velocities than the center of the curve for parallel beam elements. The solution of this problem has been discussed elsewhere.[9] In the present apparatus, the detector and source widths are sufficiently small, 2.5×10^{-3} cm, so that the widths have a negligible effect on the shape of the velocity distribution.

GENERAL DESCRIPTION OF APPARATUS

The vacuum envelope consists of the oven, rotor and detector chambers. Other than the pumping connections, through which a free flow of gas between chambers cannot occur, the only openings between the rotor chamber and the other two chambers are the slits shown in Fig. 2.

A large liquid nitrogen cooled surface is provided in each chamber, to increase the pumping speed of the cold traps for condensable gases. Pressures of 1.2×10^{-7} and 1.3×10^{-7} mm of mercury were obtainable in the detector and rotor chambers respectively, with the rotor at rest.

The velocity selector was provided with a four-to-one stepup gear built into the mount. Power was transmitted to the rotor from the outside of the vacuum envelope by a shaft which rotates in a long, lubricated phosphor bronze bearing. As shown in Fig. 2, two holes were drilled perpendicular to the axis of the bearing, one on the high pressure side so the lubricant could be replenished and one in the middle which was pumped by a mechanical pump. The vacuum seal functioned very well, though there was some tendency for the pressure to increase with increasing rotor speed. The velocity selector was seldom run over 4000 rpm, at which speed the vacuum in the rotor chamber increased to about 10^{-6} mm of mercury.

OVENS

The ovens used in these experiments were much like conventional molecular beam ovens, the main differences being in the slit construction and the oven material Since the interpretation of velocity distribution data requires that the beam be produced at a known equilibrium temperature, the ovens must be made of high conductivity material to avoid significant temperature gradients. Thus the present ovens were constructed of oxygen-free, high-conductivity copper, instead of the customary iron or nickel. Preliminary experiments in the present work have shown that in the neighborhood of 900°K, temperature differences of 30°C occur between the front and back of a conventional iron oven. This temperature difference is roughly six times the estimated accuracy of the measured temperatures. When the same measurements were repeated with a copper oven, the temperature difference was reduced to 3.5°C.

Kinetic theory considerations indicate that the width of the slit in a direction parallel to the direction of the propagation of the beam must be much smaller than the mean free paths of the atoms in order to eliminate scattering in the neighborhood of the slit. The defining slits were made of 0.0038 or 0.0025 cm thick steel strips held on the face of the oven with copper strips, whose edge extended to within about 0.25 mm of the slit itself, so that the orifice was determined by the thin steel strips. At a nominal temperature of 900°K, a thermocouple inserted in one of the copper strips indicated a temperature 0.7°C lower than the temperature measured with a thermocouple mounted in the front part of the oven as shown in Fig. 3. This temperature difference is small compared to the estimated

± 0.5 percent accuracy of the temperature measurement and has no observable effect on the results.

The oven temperatures were measured with Chromel *P*-Alumel thermocouples peened directly into the oven as shown in Fig. 3. All available information and auxiliary checks against a Pt, Pt-Rh thermocouple indicate the temperature measurement to be accurate to ± 0.5 percent of the measured temperature in degrees absolute. The beam intensity is very sensitive to the temperature of the oven, so a temperature control unit, essentially an on-off switch to control a portion of the oven heater current, was employed to maintain the oven within 0.25°C of the nominal desired temperature.

The beam was detected on a tungsten surface ionization detector. In the case of thallium, oxygen was sprayed on the detector wire to increase the detection efficiency. The filament was conditioned so that the detected beam was relatively insensitive to the filament temperature in the neighborhood of the operating temperature. The detection efficiency for potassium can be made greater than 99 percent,[11] while for thallium this may not be true. Preliminary experiments have shown that a variation of the over-all detection efficiency has no observable effect on the shape of the velocity distribution as long as the over-all detection efficiency remains constant during the run. This shows that if there is a velocity-dependent detection efficiency, it is not very sensitive to the over-all detection efficiency.

RESULTS

Differentiation of Eq. (5) shows that $\alpha^2 v_0^2$, or x_0^2 equals two at the intensity maximum, so that Eq. (5) can be written

$$I_v = 8\gamma V^4 \exp(-2V^2), \tag{8}$$

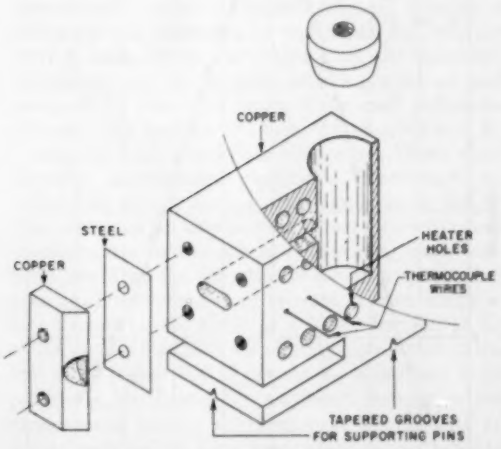


FIG. 3. Copper oven used for potassium showing details of slit construction.

[11] G. E. Cogin and G. E. Kimball, J. Chem. Phys. **16**, 1035 (1948).

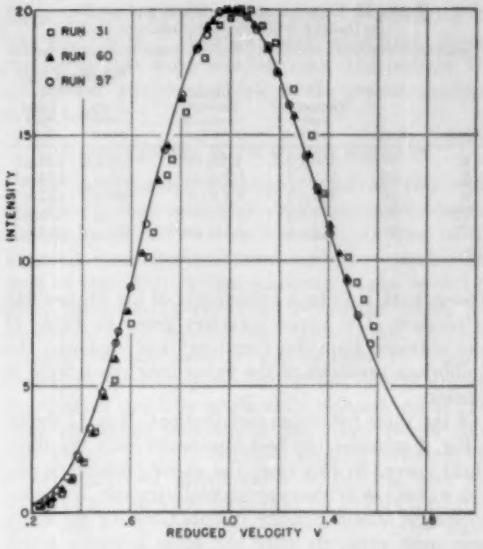


FIG. 4. Typical potassium velocity distributions. The vapor pressures in the ovens are given in Table II. Run 31 was made with thick oven slits and runs 57 and 60 with thin slits.

in which the dimensionless variable V, the reduced velocity, is the atom velocity v_0, divided by the velocity of the intensity maximum of the distribution. This is a very convenient expression for examining the agreement between theoretical and experimental curves since Eq. (8) is a theoretical "universal curve," that is, it will be the same for all velocity distributions which result from a single molecular species. To take into account the effect of finite rotor slot width, the actual theoretical distribution used in this work was a "universal curve" obtained from Eq. (6) with x_0 replaced by $V\sqrt{2}$.

Figure 4 shows universal plots of typical velocity distributions for potassium; each distribution corresponds to different experimental conditions. To compute the reduced velocity V, the experimental velocities were divided by the theoretical velocity of the intensity maximum calculated from the measured oven temperature. Since the uncertainty in temperature measurement does not exceed ± 0.5 percent, the velocity corresponding to the maximum intensity in the theoretical distribution will be accurate to ± 0.25 percent. The observed intensity measurements have been multiplied, in all cases, by an appropriate factor to give coincidence of theoretical and experimental intensities at the maximum.

Some of the important results are tabulated in Table II. The velocity at which the maximum intensity occurs can be determined directly from the experimental velocity distribution and may also be calculated from the oven temperature. The agreement of the two velocities is one criterion which must be satisfied if the

TABLE II. Experimental conditions and results for measured velocity distributions.

Beam	Run	Temperature deg K	Oven pressure in mm of mercury	Velocity of I_{max} in m/sec: From oven temp	From exp distr
K	57	466±2	4.5×10^{-3}	628±2	630±3
K	60	544±3	1.2×10^{-1}	678±3	679±3
K	31	489±2	1.4×10^{-2}	644±2	682±3
Tl	99	870±4	3.2×10^{-3}	376±1	376±2
Tl	97	944±5	2.1×10^{-2}	392±1	395±2

present work is to be a critical test of the Maxwellian distribution. The vapor pressures given in Table II were obtained from the literature[12] and represent the equilibrium pressures of the vapor over the sample in question.

Of the three potassium distributions, Run 57 shown in Fig. 4 provides the best agreement with the theoretical curve. In this case, the vapor pressure in the oven was as low as was experimentally feasible. It was not possible to obtain velocity distributions for markedly lower oven pressures since the beam intensity which depends directly on the oven pressure, would then be so low that the several sources of noise would give data of limited value long before statistical fluctuations play a significant role. The experimental points are seen to be in excellent agreement with the theoretical curve over the major part of the distribution. However, there is an observable deficiency of atoms on the low velocity side of the maximum. The value of v_{max} calculated from the oven temperature agrees with experimental value.

It has been observed that the deficiency of atoms on the low velocity side increases with increasing oven pressures and with increasing slit depth. The deficiency also increases with decreasing velocity. Run 60 was made with the same oven as Run 57 but the potassium vapor pressure was markedly increased. The experimental points of Run 60 are in good agreement with the theoretical curve in the neighborhood of the intensity maximum, but there is a pronounced deficiency of atoms on the low-velocity side and an observable excess of atoms on the high-velocity side. It appears as if the velocity distribution had undergone a velocity-dependent scattering which was not serious enough to shift the maximum of the distribution.

Run 31 illustrates the effect of a deep slit on the velocity distribution. These slits were made of copper strips which were 0.317 cm thick. In this case, the entire distribution has been shifted in the high velocity direction. It was also found that when the 0.003-cm or 0.004-cm steel shims which formed the orifice from which the beams effused from ovens like those shown in Fig. 3 were omitted, so that the orifice was determined by the knife edges of the copper strips, the velocity distributions were shifted slightly to the high velocity side of the theoretical curves. The shift corresponds to a beam temperature of the order of three percent higher than the measured oven temperature. The depth of the knife-edge copper slits was estimated to be 0.008 cm to 0.012 cm.

The potassium distributions were not extended further into the high-velocity region because this extension requires what appeared, at the time, to be a dangerously high speed of revolution. In the case of thallium, more complete distributions were obtained.

The thallium distributions are quite similar to the potassium distributions already discussed. Runs 99 and 97 shown in Fig. 5 indicate an excellent agreement between experimental points and the theoretical curves, better than previously noted. It is seen that the largest discrepancies occur on the high-velocity side of the maximum, where there is a small excess of atoms in the experimental distribution. It should be noted that the experimental points could be plotted with the high-velocity side matched to the theoretical curve. The intensities at the maximum velocity would no longer coincide and the experimental distribution would then appear to be deficient of atoms on the low-velocity side. The observed discrepancies are again more pronounced at higher oven pressures.

When thallium beams were observed, the tungsten detector wire was constantly sprayed with oxygen to maintain a good oxidized surface on the wire. The quantity of oxygen sprayed into the apparatus was kept so small that the increase in pressure due to the oxygen was about 5×10^{-3} mm of mercury in the detector chamber and less in the other chambers.

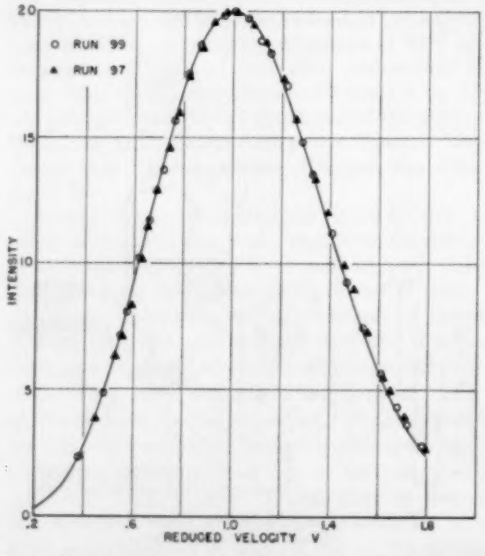


FIG. 5. Typical thallium velocity distributions. The data were taken with thin oven slits at vapor pressures given in Table II.

[12] S. Dushman, *Vacuum Technique* (John Wiley and Sons, Inc., New York, 1949), pp. 746, 748.

BEAM PURITY

The vendor of the potassium reported that it contained 2 percent sodium and 0.5 percent lithium. Because of the low vapor pressure of these two constituents compared to that of potassium, no measureable amount of sodium or lithium is expected to be present in the beam. A search was made for sodium transitions in the potassium beam with a magnetic resonance atomic beam apparatus, but no sodium transitions were observed. The sensitivity of the test was such as to make it unlikely that an amount of sodium sufficient to give an observable distortion of the velocity distribution was in the beam.

The thallium metal was reported to be 99.95 percent thallium, exclusive of the oxide on the surface. The remaining 0.05 percent was largely lead, copper, and cadmium, all of which are undetectable. Since no beam was detected with an unoxygenated detector wire, it is certain that there were no measurable amounts of cesium, rubidium, or potassium in the thallium beam.

There is the possibility that the beam may also contain dimers of the atomic species. In the case of thallium and potassium the beam in question is investigated with an atomic beam apparatus that will reject atoms with magnetic moments of the order of a Bohr magneton, which characterize the atomic states of potassium and thallium. Molecules of K_2 are in a $^1\Sigma$ states as are, presumably, those of the Tl_2 if they exist. They have magnetic moments of the order of a nuclear magneton and are not significantly deflected by inhomogeneous magnetic fields which deflect atoms through great excursions.

The molecule of potassium, K_2, is known to be present in a potassium beam, but the data on the fraction of molecules is inconsistent. In a molecular beam experiment[12] in which the nuclear magnetic moment of potassium was measured, it was found that the ion current produced at the detector by the molecular component may be as great as one percent of that produced by the total beam. Since each potassium molecule may produce two ions, the abundance of K_2 is probably a little less than 0.5 percent. Under more favorable experimental conditions, Rosenberg[14] obtained 0.25 percent K_2 in a potassium beam at 500°K. The fraction of K_2 may be shown from thermodynamic considerations to increase with increasing oven temperatures. Thus the distortion of the velocity distribution which accompanies an increase in oven temperature cannot be the result of a larger fraction of K_2 as the dimer would shift the velocity distribution to the low velocity side. If 0.25 percent K_2 is assumed to be present in the beam, the maximum increase in the experimental velocity distribution at any velocity at which observations have been made would be only 0.5 percent of the maximum intensity. The effect of K_2 is negligible in the present work. A search for Tl_2 was made with a molecular beam apparatus. It was concluded that there was less than 0.05 percent Tl_2, an amount wholly negligible in the present analysis, in the beam.

[12] Kusch, Millman, and Rabi, Phys. Rev. **55**, 1176 (1939).
[14] R. Rosenberg, Phys. Rev. **55**, 1267 (1939).

COMPARISON WITH OTHER RESULTS

The experimental techniques used in two early attempts to measure velocity distributions[1,2] were not sufficiently refined to permit an accurate test of the Maxwell velocity distribution law to be made. The work of both Eldridge and Lammert shows deviations from theoretical predictions, deviations which are greater than the experimental uncertainties. Zartman[3] investigated the velocity distributions in bismuth beams. Unfortunately, no critical test of the Maxwell distribution could be made since bismuth vapor contains significant amounts of molecular species other than atomic Bi. Thus Zartman actually measured the superpositions of distributions of Bi, Bi_2, and Bi_3. When he adjusted the amount of Bi_2 so that the agreement was good on the low-velocity side, there was a pronounced excess of molecules on the high-velocity side. The amount of Bi_3 was small so that it had little effect on the major part of the distribution.

Ko[4] used essentially the same apparatus as Zartman to measure the fraction of Bi_2 in bismuth beams as a function of the oven temperature and pressure. From his data, he was able to calculate the dissociation energy of the Bi_2 molecule. When the amount of dimerization was adjusted to give good agreement on the low-velocity side of the distribution, there were too many molecules on the high-velocity side. He also found evidence that Bi_3 was present in the beam.

Cohen and Ellett[5] measured velocity distributions in sodium and potassium beams. As a velocity selector they used an inhomogeneous magnetic field which was perpendicular to the direction of propagation of the beam. Since the deflection of beam atoms is proportional to the gradient of the magnetic field, it is important that the field inhomogeneity be constant over the height and width of the beam. They assumed the gradient to be constant if the observed distribution could be fitted to the theoretical distribution after the adjustment of constants not determinable from *a priori* considerations. The nature of the constants is such that an independent check of the observed maximum and that deduced from the temperature cannot be made. Subject to these limitations, they observed no systematic deviation from Maxwell's theory at low oven pressures. At high oven pressures, they observed serious deviations from theory, notably a deficiency of atoms on the low-velocity side. They predicted that knife-edge slits would give better results than their 2-mm thick rectangular slits. However, on the basis of the present work, there is some doubt that the theoretical distributions could be observed with deep slits even at low pressures.

In 1947, Estermann, Simpson, and Stern[6] measured the velocity distributions in cesium beams by use of the free fall due to gravity as a velocity selector. They observed a pronounced deficiency of atoms on the low-velocity side and noted that the discrepancy increased with increasing oven pressure. They attributed the deficiency to collisions near the oven slit. An order of magnitude calculation is given which indicates that the deficiency might be explained on the basis of Cs-Cs collisions in the neighborhood of the slit.

The most recently reported results on velocity distributions are those of Bennett and Estermann[9] who measured potassium velocity distributions. Their experimental velocity distributions were in good agreement with theoretical distributions on the high-velocity side of the intensity maxima, but there was always a deficiency of atoms on the low-velocity side, a deficiency which increased with increasing oven pressure and with a decreasing velocity. Even at the lowest oven pressure at which experimental results were obtained, 1.8×10^{-3} mm of mercury, the intensity was about 90 percent of the theoretical intensity for atoms at a velocity one-half the velocity of the intensity maximum. The deficiency of atoms on the low-velocity side was greater if the oven slits were rectangular than if they were knife edges. The velocities at which the experimental intensity maxima occurred were usually found to agree within three percent with the predicted values.

In the present work, experimental velocity distributions of both potassium and thallium were found to be in good agreement with theoretical velocity distributions. It has been shown that the experimental and theoretical distributions, for small beams, do not differ by more than one percent of the maximum intensity over the entire range of the measured distributions, extending from as low as 0.2 to as high as 1.8 times the velocity of the intensity maximum. Furthermore, the velocities at which the intensity maxima occur agree within 0.5 percent with velocities calculated from the measured oven temperature. These results provided the best agreement to date between observed distributions and theoretical distributions calculated from the assumption that the beam is a consequence of a Maxwellian distribution within the oven which effuses through an ideal slit.

REFLECTION FROM THE ROTOR SLOTS

It has been assumed that all molecules which strike the wall of the slots in the velocity selector are removed from the beam. If the beam were very narrowly collimated so that the detector would intercept only atoms which arrive along a line parallel to the rotor axis and which leave the oven along the same line, the assumption would be justified. However, no collimation other than that provided by a source and detector of small width was provided in the original design of the apparatus. An attempt was made to improve the collimation of the beam by placing a 0.005-cm slit at each end of the velocity selector; but it was impossible to observe a velocity distribution with this arrangement, since vibration caused by the rotating velocity selector produced serious unsteadiness in the beam. Sufficient enlargement of the collimation slits to reduce variations in the beam intensity consequent to vibration reduces the collimation of the beam to the point where the slits no longer serve to prevent the detection of atoms which may have undergone reflections from the wall of the slots and the slits were eventually removed. However, before the slits were removed, an attempt was made to reflect potassium atoms from the wall of one of the two straight slots when the rotor was at rest. There was no evidence at all of specular reflection. The diffused reflection and scattering of atoms was so small that the intensity intercepted on the detector when not in the direct line of the beam was less than 0.1 percent of the beam and hence not important in the present work.

SCATTERING FROM RESIDUAL GAS

The theoretical velocity distribution to be observed at the detector has been derived under the assumption that a Maxwellian distribution occurs in the oven and that the slit of the oven is an ideal aperture. It is interesting to investigate the occurrence of a possible distortion of the velocity distribution due to velocity dependent scattering of the beam atoms by the residual gas in the apparatus. The mean free path of a molecule with velocity v in a single Maxwellian gas is[15]

$$L_v=\alpha_1 v^4/n_1 S \times\left[\exp(-x^2)+(2x+x^{-1})\int_0^x \exp(-y^2)dy\right]^{-1}, \quad (9)$$

where $\alpha_1^2=(1/2)m_1/(kT_1)$ in which the subscript 1 refers to the scattering gas, n_1 is the density of the gas, x is equal to $\alpha_1 v$ and S is the collision cross-sectional area. To make any calculations from Eq. (9), one has to assign a value to α_1 and the cross-sectional scattering area. In the case at hand, the scattering gas probably consists of air plus all sorts of condensable vapors, including pump oil dissociation products. The temperature of the residual gases is not defined but an "effective" temperature may range from slightly above room temperature to liquid nitrogen temperatures. Since the vacuum in the apparatus tends to deteriorate with increasing rotor speed, there may be a change in compostion of the gas at different values of x_0. Thus it is virtually impossible to calculate L_v for this situation. To estimate the magnitude of the effects due to scattering, it will be assumed that the residual gas approximates air at room temperature and that the cross-sectional scattering area is 4.0×10^{-14} cm^2. This area is

[15] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938), p. 109.

close to the value found by Mais[16] and Rosenberg[14] for the scattering of potassium beams by nitrogen.

The probability that an atom will travel from the oven slit to the detector wire without a collision is

$$P_v = \exp(-s/L_v), \tag{10}$$

where s is equal to the distance between the oven slit and the detector wire, 60 cm. Then P_v can be calculated from Eqs. (9) and (10). If Eq. (5) is multiplied by the appropriate value of P_v when account is taken of the variation of the pressure with rotor velocity, a distribution is obtained whose shape is only trivially distorted on the scale of Figs. 4 and 5. In view of the uncertainty in the composition of the gas, the gas temperature, and the related cross sections, the result is subject to a corresponding uncertainty. However, the excellent results for low and moderate beams of potassium and thallium make it clear that any distortion due to scattering by the residual gas in the apparatus must be small.

DISCUSSION AND CONCLUSIONS

The results presented in this work demonstrate that the experimental velocity distribution can be predicted with good accuracy from the assumptions that the velocity spectrum in the oven is Maxwellian and that effusion occurs through an ideal aperture. It may be thought that the excellent agreement is somewhat fortuitous, i.e., that an oven slit design was discovered with the property that a non-Maxwellian distribution effusing through a nonideal aperture gives a distribution experimentally indistinguishable from that arising under our assumptions. An ideal aperture is one in which the dimensions of the aperture are much smaller than the mean free path of the molecules in the gas. The slit used in the present experiments has been designed precisely to meet this condition within the range of experimental feasibility. The agreement between theoretical prediction and observed distributions is therefore, not fortuitious and the result of the present work demonstrates within relatively narrow limits of error that a gas in thermal equilibrium with its surroundings possesses a Maxwellian velocity distribution.

In view of the success of this work, a study of the velocity distributions of alkali halide molecular beams has been made, in which it has been assumed that the experimental velocity spectrum is the superposition of the Maxwellian velocity distribution of the molecular components which are in thermal and chemical equilibrium within the oven chamber. The results of this work will be published elsewhere.

[16] W. H. Mais, Phys. Rev. 45, 773 (1934).

ACKNOWLEDGMENTS

The authors wish to thank Professor I. I. Rabi for many helpful discussions. The authors also take this opportunity to express their appreciation to Mr. Joseph Haas of the Columbia Radiation Laboratory for the construction of the cylindrical velocity selector with the 702 perfectly milled slots.

Spin Resonance in Metals as a Function of the Overhauser Effect

JEROME I. KAPLAN
University of California, Berkeley, California, and Naval Research Laboratory, Washington, D. C.
(Received April 22, 1955)

The polarization of the nuclear spin states by saturating the electron spin levels as predicted by Overhauser is shown to alter the electron spin resonance absorption line shape. It is shown that under certain stringent conditions the line shape may depend upon the sweep direction of the magnetic field H_0. The theory is developed under the assumption of no skin-depth broadening.

IT will be shown that under certain conditions the Overhauser[1] effect causes the electron spin resonance absorption line shape in metal particles, small compared to the skin depth, to be dependent on the sweep direction of H_0, the applied static field. The rate of energy absorption P for the spin system, assuming no skin-depth broadening,[2] is given by[3]

$$P=\frac{\omega\hbar}{2H_1^2}\frac{\alpha^2 g(\omega-\omega_0')n_0}{1+\alpha^2 g(\omega-\omega_0')}=\frac{\omega\hbar}{2H_1^2}n_0 s, \tag{1}$$

where $\alpha^2=(\pi/4)\gamma_e^2 H_1^2 T_1$ and the saturation factor s is

$$s=\frac{\alpha^2 g(\omega-\omega_0')}{1+\alpha^2 g(\omega-\omega_0')}.$$

ω is the applied frequency, H_1 the rf magnetic field, T_1 the spin lattice relaxation time, ω_0' the electronic Larmor frequency, γ_e the electronic magnetomechanical ratio, n_0 the population difference of the electronic spin states as determined by the total static field H_0' and $g(\omega-\omega_0')$ the normalized shape factor for the transition. For a Fermi gas of electrons, the population difference is

$$n_0=3N_e\hbar\gamma_e H_0'/4E_f, \tag{2}$$

where E_f is the Fermi energy and N_e the electron density.

It is assumed that the line-shape function $g(\omega-\omega_0')$ has a Lorentzian form which expressed in terms of magnetic fields becomes

$$g(\omega-\omega_0')=\frac{(\pi\gamma_e H_T)^{-1}}{1+H_T^{-2}(H-H_0')^2}, \tag{3}$$

where $\omega=\gamma_e H$ and $1/T_1=\gamma_e H_T$. Substitution of $g(\omega-\omega_0')$ in the expression for the saturation factor s gives the result that

$$s=\beta^2/[(1+\beta^2)+H_T^{-2}(H-H_0')^2], \tag{4}$$

where

$$\beta^2=\alpha^2/\pi\gamma_e H_T.$$

The static field acting on the electron is the sum of the applied field and the contribution arising from the polarization of the nuclei. According to Overhauser, H_0' is given as

$$H_0'=H_0+\mathcal{P}\Delta H, \tag{5}$$

where

$$\Delta H=(8\pi/3)N|\psi_0^2|I$$

is the maximum local field and

$$\mathcal{P}=\sum_{m=-I}^{m=I} m\exp(m\hbar\gamma_{eff}/2)\Big/\left(I\sum_{m=-I}^{m=I}\exp(m\hbar\gamma_{eff}/2)\right) \tag{6}$$

is the measure of nuclear polarization.

$$\gamma_{eff}=(\gamma_e+s\gamma_n)H_0/kT,$$

N is the density of atoms, ψ_0 the value of the electronic wave function at the nucleus, I the nuclear spin, γ_n the nuclear magnetomechanical ratio, k the Boltzmann constant, and T the temperature. For $I=\frac{1}{2}$, $s\gamma_e\gg\gamma_n$, and $\hbar s\gamma_e H_0\ll kT$,

$$H_0'-H_0\simeq\tfrac{1}{2}(s\hbar\gamma_e H_0)\Delta H/kT. \tag{7}$$

The saturation factor s in Eq. (7) is replaced by its value given in Eq. (1). The resulting equation,

$$H_0'-H_0\simeq\frac{\hbar\gamma_e\Delta H(1+\beta^2)^{-1}\beta^2}{2kT}\times\frac{H_0}{1+H_T^{-2}(1+\beta^2)^{-1}(H-H_0')^2}, \tag{8}$$

relates H_0' to H_0, the other factors being taken as constants.

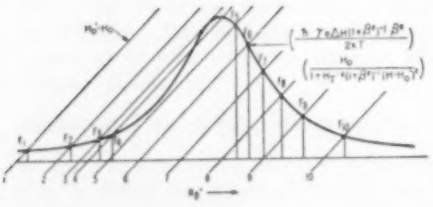


FIG. 1. Graphical solution of the equation

$$H_0'-H_0=\hbar\gamma_e\Delta H(1+\beta^2)^{-1}\beta^2\times\frac{H_0}{1+H_T^{-2}(1+\beta^2)^{-1}(H-H_0')^2}$$

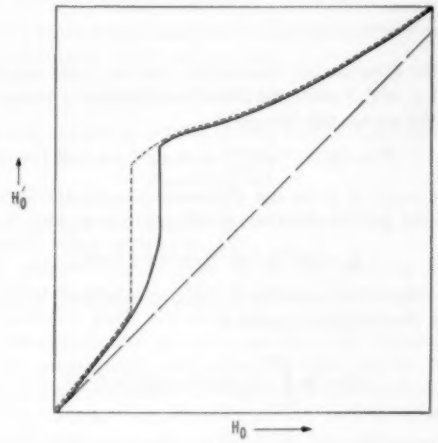
where it is assumed that $\hbar\gamma_e\Delta H(1+\beta^2)^{-1}\beta^2 H_0=4/3$ and $H_T^{-2}\times(1+\beta^2)^{-1}=1$.

[1] A. Overhauser, Phys. Rev. **92**, 411 (1953).
[2] F. Dyson, Phys. Rev. **98**, 439 (1955).
[3] A. Portis, Phys. Rev. **91**, 1071 (1953).

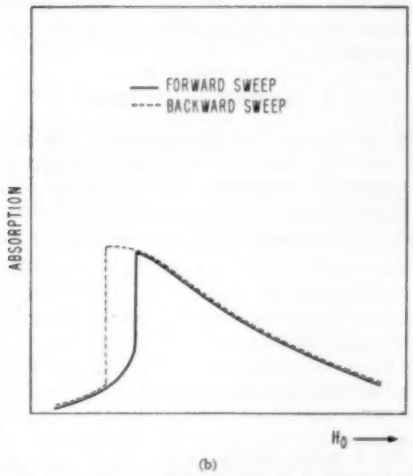
The energy absorption in terms of H_0 and H_0',

$$P=\frac{\omega\hbar}{2H_1^2}\left(\frac{(2kT)}{\gamma_e\Delta H}\right)\left(\frac{3N_e\gamma_e\hbar}{4E_f}\right)\left(\frac{H_0'}{H_0}\right)(H_0'-H_0), \quad (9)$$

is found by using Eqs. (1), (2), and (7).



(a)



(b)

FIG. 2. (a) Plot of H_0' vs H_0 as obtained from Fig. 1. (b) Plot of the energy absorption vs H_0 as obtained from Fig. 2(a) and Eq. 9.

For each value of H_0, H_0' is calculated from Eq. (8). The qualitative form of the solution can be found graphically as shown in Fig. 1. As H_0 is increased from zero through resonance the solutions for H_0' are marked by F. It is to be noted that the height of the Lorentz curve in Fig. 1 will vary as H_0, but this will only change the magnitude and not the form of the solution. The nteresting point is that there can be a discontinuity in

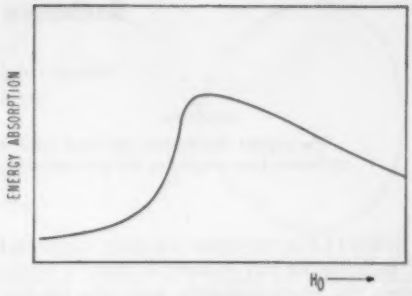


FIG. 3. Plot of the energy absorption vs H_0 where H_T is assumed double that used in Fig. 1.

H_0' as H_0 varies from position 4 to 5. The plot of H_0 vs H_0' for forward and backward sweeps as found from Fig. 1 is shown in Fig. 2(a). The substitution of these values in Eq. (9) gives the form of the absorption curve for both sweep directions as shown in Fig. 2(b). In Fig. 3 is shown a plot of the absorption for the case where H_T has been increased by a factor of two over the value assumed in Fig. 1. The effect of this change is to broaden out the Lorentzian curve in Fig. 1 also by a factor of two. The broadened curve intersects the straight line $H_0'-H_0$ only at one point, corresponding to a single solution of H_0' as a function of H_0. For this case then the absorption will be independent of the direction of sweep. It is seen that a necessary condition for the asymmetry as regards direction of sweep is that the slope at the inflection point of the Lorentz curve in Fig. 1 be greater than unity. In terms of the physical constants of the problem and the value of H_0 in the resonance region, this condition requires

$$\frac{\hbar\gamma_e^2T_1H_1^2\Delta H\sqrt{3}}{8kTH_T^2}\times\frac{H_0}{[(1+\gamma_e^2T_1H_1^2)/4H_T]^{\frac{1}{2}}}>1. \quad (10)$$

Thus, if $H_0=10^3$, $\Delta H=10^2$, $H_1=1$, $8kT/\sqrt{3}=10^{-15}$, $H_T=1$, $\gamma_e=10^7$, $T_1=10^{-7}$, and $\hbar=10^{-27}$, the expression on the left side of the inequality in Eq. (10) will approximately equal one.

The value of dP/dH_0, to the approximation where terms in $[(H_0'-H_0)/H_0]^2$ are neglected, can be written as

$$\frac{dP}{dH_0}\simeq\frac{\omega\hbar}{2H_1^2}\left(\frac{3N_e\gamma_e\hbar}{4E_f}\right)\left(\frac{2kT}{\gamma_e\Delta H}\right)\times\left(1+\frac{2(H_0'-H_0)}{H_0}\right)\frac{d[H_0'-H_0]}{dH_0}. \quad (11)$$

ACKNOWLEDGMENTS

I wish to thank Professor C. Kittel for suggesting this problem and the Office of Naval Research for supporting the research at the University of California.

Radiations from a Spinning Rod

DAVID PARK*
Williams College, Williamstown, Massachusetts
(Received March 17, 1955)

The angular distribution and total value of the power carried in the gravitational and electromagnetic radiations from a spinning rod are evaluated and compared.

RECENTLY in the course of a study of gravitational radiation we had occasion to make a comparison between the electromagnetic and the gravitational radiation of energy from a spinning rod—the only example which seems at all simple to work out in the gravitational case.[1] Since the results of the comparison are by no means obvious in advance, we shall give a brief account of them here.

1. Gravitational Theory

Let J_{ik} be the tensor which gives the moment of inertia of the rod,

$$J_{ik}=\int(3x_ix_k-\delta_{ik}r^2)\rho(\mathbf{r})dv \quad (x_i,x_k=x,y,z), \tag{1}$$

where $\rho(\mathbf{r})$ is the mass density. It has been shown from the general theory of relativity that if the rod is long and narrow, and its ends are moving at a speed much less than that of light, and if $\mathbf{n}$ is a unit vector in the direction of observation, then the power per unit solid angle radiated in this direction is[2]

$$P_n(G/32\pi c^5)[\tfrac{1}{4}(n_in_kA_{ik})^2+\tfrac{1}{2}A_{ik}^2-n_in_kA_{ij}A_{jk}]_{\text{ret}}, \tag{2}$$

where

$$A_{ik}=d^2J_{ik}/dt^2. \tag{3}$$

To evaluate this we can choose coordinates such that the rod spins about the z-axis and at the retarded time lies instantaneously along the x-axis. At this instant only one independent component of J survives, and

$$A_{xy}=A_{yx}=-12I\omega^2, \tag{4}$$

where I is the rod's moment of inertia about the z-axis, or $\int\rho r^2dv$. With this, (2) is

$$P_n=(4GI^2\omega^6/\pi c^5)[(n_xn_y)^2+n_x^2].$$

If the direction of observation has the polar angles θ and φ, with θ measured from the z-axis and φ vanishing on the x-axis, this becomes[2]

$$P_n=(4GI^2\omega^6/\pi c^5)[\tfrac{1}{4}\sin^4\theta\sin^2 2\varphi+\cos^2\theta].$$

The angle φ gives the difference in azimuth between the rod and the observer. Averaging over φ gives

$$P_n=(4GI^2\omega^6/\pi c^5)[\cos^2\theta+\tfrac{1}{8}\sin^4\theta], \tag{5}$$

and this is the radiation pattern, as illustrated in Fig. 1. The total radiated power is

$$P_T=2\pi\int_0^\pi P_n\sin\theta d\theta=32GI^2\omega^6/5c^5 \tag{6}$$

in agreement with Eddington's result.[1] As a result of this energy loss the rod will spin more and more slowly, and the radiated energy, like that in Maxwell's theory, will either impart motion to surrounding objects or be lost in outer space.

2. Electromagnetic Theory

The radiation in the electromagnetic case is somewhat arbitrary, since the rod's center of charge does not necessarily correspond with its center of mass. But in order to get results comparable with those just found, let us suppose that it does, so that the electric

FIG. 1. Angular distribution of the intensity of the gravitational power radiated by a rod spinning about a vertical axis through its center of mass.

* Visiting Fulbright lecturer in physics, University of Ceylon, 1955–1956.

[1] The problem of determining the total power radiated was first studied by Einstein [A. Einstein, Sitzber. preuss. Akad. Wiss., Physik-math. Kl. 1916, p. 688]. A computational error led to a second attack (*ibid.*, 1918, p. 154). That here also there is a mistake [it is a factor of 2 in Eq. (16) of Einstein's second paper] was pointed out by Eddington in his discussion [Proc. Roy. Soc. (London) **A102**, 268 (1922); *The Mathematical Theory of Relativity* (Cambridge University Press, Cambridge, 1924), second edition, Note 8]. The question of the radiation pattern does not seem to have been discussed before.

[2] For convenience, we use the form given by Landau and Lifschitz *The Classical Theory of Fields* (Addison-Wesley Press, Cambridge, 1951), p. 331.

[3] This formula follows also from Eq. (18) of Eddington's paper cited in footnote 1.

dipole moment of the rod is zero and the magnetic dipole moment is constant. The radiation is therefore electric quadrupole, and using the notation of (1) and (3), but with ρ now denoting the charge density, we have for the Poynting flux per unit solid angle in the direction $\mathbf{n}$,[4]

$$P_{\mathbf{n}}=(1/144\pi c^5)[n_i n_k A_{ij} A_{jk}-(n_i n_k A_{ik})^2]. \tag{7}$$

If we introduce polar angles as before and average over all orientations of the rod, we find

$$P_{\mathbf{n}}=(I^2\omega^6/2\pi c^5)(1-\cos^4\theta), \tag{8}$$

and this pattern is plotted in Fig. 2. The total power given out is

$$P_T=8I^2\omega^6/5c^5, \tag{9}$$

where I is, of course, the electrostatic analog of the quantity which appears in (5).

In order to understand better the striking dissimilarity between the two patterns, let us consider the radiation along the rod's axis of rotation. In the electromagnetic case, symmetry excludes any transverse electric or magnetic field along the axis, but for the gravitational potentials the situation is quite different. Writing g_{ik} as $\eta_{ik}+h_{ik}$, where η_{ik} are the Galilean values, we find that only the transverse components h_{xx} h_{yy}, and h_{xy} do not vanish on the axis, and it is well known that these components carry energy. If one evaluates them in a new coordinate system whose z-axis coincides with the old one and whose x- and y-axes are rotated so as to make h_{xy} vanish, we find that at any point z this system is rotating at a rate 2ω, its angular position being given by $2\omega(t-z/c)$. The gravitational field spirals out along the axis, so to speak, with the speed of light.

In order to clarify further the question of the angular distribution of these two sorts of quadrupole radiation, let us consider the following idealized situation. There are two masses, M and m, with $M\ll m$, whose center of gravity is at rest at the origin. By means of an essentially massless arrangement of rods and ropes some people in the large mass impart to the small one a velocity $\mathbf{v}(\ll c)$ and a uniform acceleration $\mathbf{a}$. We shall find the quadrupole gravitational radiation due to this motion and, assuming that the charges are in the same ratio as the masses, the quadrupole electromagnetic radiation also. The inertia tensor is

$$J_{ik}=m(3x_i x_k-\delta_{ik}r^2)$$

plus a corresponding term contributed by M. But since x_i and its derivatives vary inversely with the masses, and since they appear quadratically in J_{ik}, the latter contribution can be neglected. Evaluation of (2) and (7) now gives

$$P_{\mathbf{n}}=(81Gm^2/32\pi c^5)[\mathbf{a}\times\mathbf{n}]^2[\mathbf{v}\times\mathbf{n}]^2, \tag{10}$$

with

$$P_T=(27Gm^2/20c^5)[3a^2v^2+(\mathbf{a}\cdot\mathbf{v})^2], \tag{11}$$

and

$$P_{\mathbf{n}}=(9e^2/16\pi c^5)[(\mathbf{a}\cdot\mathbf{n})[\mathbf{v}\times\mathbf{n}]+(\mathbf{v}\cdot\mathbf{n})[\mathbf{a}\times\mathbf{n}]]^2, \tag{12}$$

with

$$P_T=(6e^2/5c^5)[2a^2v^2-(\mathbf{a}\cdot\mathbf{v})^2], \tag{13}$$

respectively, from which one sees that if $\mathbf{n}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{v}$, the first expression for $P_{\mathbf{n}}$ is a maximum and the second is zero, whereas if $\mathbf{n}$ is parallel to $\mathbf{a}$ or $\mathbf{v}$, the reverse is true.

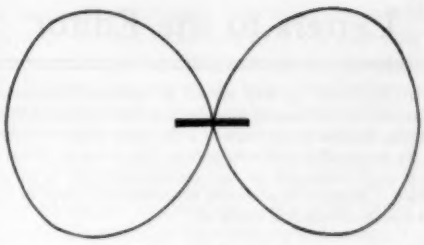


FIG. 2. Angular distribution of the intensity of the electromagnetic power radiated by a charged rod spinning about a vertical axis through its center of charge.

This difference in patterns is characteristic of the nature of the field quantities. It does not reflect the nonlocalizability of energy in a gravitational field, for though the density of this energy is indeed not an observable quantity, the flux which we have calculated can in principle be measured. (We have verified this conclusion by computing the energy-flux terms in Pauli's modified expression for the gravitational energy[5] and showing that they reduce to the expressions given above.) And whereas it is natural to attempt to account for the difference in the numerical coefficients in (6) and (9) by the fact that the gravitational field has the more components, it can unfortunately be shown that the number of independent components is only two in each case.[6]

This investigation was carried out in a seminar in cooperation with L. Scarfone and W. Williams.

[4] See reference 2, p. 206.

[5] W. Pauli, Physik. Z. **20**, 25 (1919).

[6] M. Fierz, Helv. Phys. Acta **12**, 3 (1939); **13**, 45 (1940); J. S. de Wet, Phys. Rev. **58**, 236 (1940).

Letters to the Editor

PUBLICATION of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is five weeks prior to the date of issue. No proof will be sent to the authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not exceed 600 words in length and should be submitted in duplicate.

Number of Vacancies Created by Heavy Corpuscular Radiation

J. NEUFELD, *Oak Ridge National Laboratory, Oak Ridge, Tennessee*
AND
W. S. SNYDER, *The University of Tennessee, Knoxville, Tennessee and Oak Ridge National Laboratory, Oak Ridge, Tennessee*
(Received June 16, 1955)

KINCHIN and Pease[1] have suggested that a moving atom which collides with and displaces another atom may, in certain cases, take its place. Thus the number of displaced atoms is not the same as the number of vacancies, even when diffusion of interstitial atoms is neglected.

Let E be the energy of the striking atom before a collision and $E-y$ its energy after the collision. Altering the assumptions of Kinchin and Pease, we assume the striking atom replaces the struck atom in the monoatomic case whenever $E-y<\alpha$ and $y>\alpha$. The parameter α is the displacement energy and is often taken as about 25 ev. Thus a vacancy is created only when $y>\alpha$ and $E-y>\alpha$.

If $g(E)$ equals the mean number of vacancies created by the displacement of an atom of the medium by ejecting an atom with energy after ejection of E, then $g(E)=1$ for $0<E<2\alpha$. For $E>2\alpha$, $g(E)$ satisfies the following integral equation:

$$g(E)=\int_0^{E-\alpha} dyK(E,y)g(E-y) + \int_\alpha^E dyK(E,y)g(y-\alpha), \quad (1)$$

where $K(E,y)dy$ is the probability that a striking atom with energy E will lose energy y in dy in its next collision. The derivation of this equation follows closely the similar discussion of Eq. (7) of the authors' paper[2] (hereafter indicated by D), and assumes that ionization can be neglected. In accordance with the theory given in D it should apply only for $E<Me^4/2\hbar^2$.

For low energies "hard sphere" scattering is assumed and $K(E,y)$ is taken as $1/E$. By the methods developed in D it is shown that

$$[(E+\alpha)/3\alpha]<g(E)<1.056[(E+\alpha)/3\alpha] \quad (2)$$

for all monoatomic substances and for $2\alpha<E<\beta$. The value β indicates the energy above which Rutherford scattering is assumed to apply and is given by $\beta=2Z^2e^2/a$ with a the Bohr radius of the atom, (D§3). For Rutherford scattering $K(E,y)=\beta^2/4Ey^2$ for $E\beta^2(4E^2+\beta^2)<y<E$ and $K(E,y)=0$ otherwise. It is established that the inequality (2) is valid for all $Z>10$ for the region of Rutherford scattering, and with minor modifications, for lower Z.

The results for $E>Me^4/2\hbar^2$ are not complete but indicate a less rapid rate of increase for $g(E)$ due to ionization.

Thus the number of vacancies seems to be about $\frac{2}{3}$ the number of displacements as calculated in D.

Harrison and Seitz[3] have calculated the change in resistivity of Cu irradiated by 17-Mev deuterons. Using the number of displaced atoms as equal to the number of vacancies they find the calculated value 5 times the observed value. The calculation of vacancies indicated above reduces this to 3.3 times the observed value.

[1] G. H. Kinchin and R. S. Pease, J. Nuclear Energy **1**, No. 3, 200–202 (1955).
[2] W. S. Snyder and J. Neufeld, Phys. Rev. **97**, 1637–1646 (1955).
[3] W. A. Harrison and F. Seitz, Phys. Rev. **98**, 1530(A) (1955).

Dipole Fields and the Helium Film

N. G. MCCRUM AND J. C. EISENSTEIN
Department of Physics, Pennsylvania State University, University Park, Pennsylvania
(Received June 6, 1955)

IN measurements of the thickness and the flow rate of liquid helium films the most frequently reproduced characteristic has been lack of reproducibility. The scatter of the values for flow rates obtained in one laboratory is not significantly less than the discrepancies existing between different laboratories.[1,2] An adequate theory of the liquid helium film must account for this salient experimental fact.

The van der Waals theory of the film[3,4] is based on very simple assumptions about the surface and the interaction of helium atoms with surface atoms. According to the theory a film several hundred angstroms in thickness exists both below and above the λ temperature. The essential validity of this theory has been established by the experiments of Ham and Jackson.[5] The theory can be improved by making physically plausible assumptions about the surface. The effect of these assumptions is to introduce into the theory a real surface in place of an idealized and nearly featureless substrate.

On most surfaces liquid will be held in irregularities by surface tension. The volume of this adsorbed capillary liquid will depend on the roughness of the surface and will be increased if impurities are present.[6]

If we make the usual assumption that the flow rate is proportional to the film cross section then it is to be expected that dirty surfaces will show large flow rates and a pronounced height dependence.[6] To distinguish it from the capillary liquid, we shall call the helium held on flat regions of the surface by direct van der Waals interaction with the substrate the film liquid. When only rather small numbers of impurity molecules are present the local electric fields above the surface will be modified and will, under suitable conditions, have an important effect on the volume of the film liquid.

When an electrostatic field E is present the potential equation which determines the film profile must be modified by inclusion of an electrostatic energy term. One then finds that the film thickness x at a distance y above the liquid surface is given by

$$x=[k/(Mgy-\tfrac{1}{2}N\alpha E^2)]^{\frac{1}{3}},$$

where k is a constant dependent on the van der Waals interaction of a helium atom with a wall atom, M the gram molecular weight of helium, g the acceleration due to gravity, N Avogadro's number, and α the polarizability of helium. By inserting $\alpha=2.10\times10^{-25}$ [7] and $k=26\times10^{-14}$ (helium on glass)[4] one finds that the film thickness 1 cm above the liquid surface will be increased 7% by a field of 100 esu and 40% by a field of 200 esu. Intense fields of the order of 10^4 esu are known to exist at the surface of a metal.[8,9] They are due to the overlap of the electron wave functions beyond the metal surface. These fields are of short range, but are sufficient to polarize the first layers of adsorbate[10] or to align permanent dipoles.

The potential at a point Q above an electric double layer is $P\omega$ where P is the dipole moment per unit area and ω is the solid angle subtended at Q by the boundary curve of the double layer. If Q is very close to the double layer then the potential will be very nearly constant and the field small. Consequently in the absence of impurities a uniform layer of induced or permanent dipoles will not appreciably affect the film thickness. However, if small amounts of impurities are adsorbed so that the surface is heterogeneous the situation is radically different. For simplicity assume that there is a circular disk of aligned dipoles on the surface. Near the center of this disk the field will be of the order of $2\pi p n b/c$ where p is the dipole moment of a single molecule, n is the number of dipoles per unit volume, b is the thickness of the disk, and c is its radius. For a disk consisting of water molecules the dipole field is of the order of $10^5(b/c)$ esu. Consequently, even for unfavorably small values of (b/c) or imperfect alignment of the dipoles, the field is still significant. The experiments of Mignolet[10] afford presumptive evidence that even nonpolar impurity would cause appreciable modification of the surface fields. A detailed discussion of the effects of impurity in film experiments will be presented elsewhere.

It would be extremely interesting to investigate adsorption, the saturated film thickness and flow rate over various surfaces in the presence of external electric fields. Barium titanate would be an ideal substance to use because it can be obtained as a single crystal or in powdered form and has a dielectric constant of about 350 at helium temperatures.[11] The effect of electric fields on adsorption would be more conveniently studied at higher temperatures with gases other than helium. The method of Garbatski and Folman,[12] using using water as an adsorbent, would be particularly suitable.

[1] B. N. Yeselson and B. G. Lazarev, Zhur. Eksptl. i Teort. Fiz. **23**, 552 (1952).
[2] B. Smith and H. A. Boorse, Phys. Rev. **98**, 328 (1955).
[3] J. Frenkel, J. Phys. U.S.S.R. **2**, 365 (1940).
[4] L. I. Schiff, Phys. Rev. **59**, 838 (1941).
[5] A. C. Ham and L. C. Jackson, Phil. Mag. **45**, 1084 (1954).
[6] R. Bowers, and K. Mendelssohn, Proc. Phys. Soc. (London) **A63**, 1318 (1950).
[7] J. C. Slater and J. G. Kirkwood, Phys. Rev. **37**, 682 (1931).
[8] J. Bardeen, Phys. Rev. **49**, 653 (1936).
[9] K. Huang and G. Wyllie, Proc. Phys. Soc. (London) **A62**, 180 (1949).
[10] J. C. P. Mignolet, Discussions Faraday Soc. No. **8**, 105 (1950).
[11] W. P. Mason and B. T. Matthias, Phys. Rev. **74**, 1622 (1948).
[12] U. Garbatski and M. Folman, J. Chem. Phys. **22**, 2086 (1954).

Andronikashvilli Disk Measurements in Liquid Helium-3—Helium-4 Mixtures*

JOHN R. PELLAM
California Institute of Technology, Pasadena, California
(Received June 23, 1955)

DIRECT measurements of the effective "normal fluid" density in a 3.3% (by mass) solution of helium-3 in helium-4 have been obtained from the λ point down to temperatures of about 0.95°K, using the Andronikashvilli[1] disk technique. Such measurements are just within the limits of experimental observation with the amounts of helium-3 (40 cc STP) available.[2] A set of 15-mica disks, 4 mm in radius and spaced 0.019 cm apart, were mounted on a vertical shaft and suspended as an Andronikashvilli disk assembly within a cylindrical cavity of about $\frac{3}{4}$ cc total volume. Quartz fiber suspensions were used, most of the measurements being taken with fibers giving periods in vacuum of 3.589 sec and 5.220 sec, respectively.

The operation of the system was such that corrections for skin-depth effects were kept comparatively small. Near the λ point the compensating effects of incomplete motion of excitations between the disks and excess motion outside the disks partly cancelled each other. Whereas for pure helium-4 the corrections would become very large at the lowest temperatures (comparable to the quantities measured), this was not the case for the mixture. The comparatively high lower limit of "normal fluid" density in the mixture, com-

TABLE I. Corrections (percentage) for skin-depth effects.

Temperature (°K)	Correction for skin depth
1.4 (and lower)	0
1.5	0.5%
1.6	0.8%
1.7	2.3%
1.8	4.0%
1.9	4.3%
2.0	3.2%
2.1	1.5%
λ point	0

bined with, apparently, a relatively low viscosity at temperatures in the region of 1°K precluded unmanageably large skin depths.

Portions of the density data are given in Fig. 1. Uncorrected data are indicated by the points and the solid curve ——— drawn through them. The observed effective mass densities of the excitations (rotons, phonons, plus helium-3) have been normalized to unity for the 3.3% mixture immediately above the λ point on the basis of an "ideal" (no correction) disk system; and are expressed in terms of the customary "normal fluid" concentration (ρ_n/ρ). The orders-of-magnitude of correction for skin-depth effects are indicated in Table I. These values were deduced from observations of oscillational damping of the disk system in the mixture. Besides the generally small extent of corrections needed, it should be noted that below about 1.5°K they become unimportant. The broken curve — — — of Fig. 1 represents the corrected value of the observed "normal fluid" concentration (ρ_n/ρ).

The most significant feature of these (ρ_n/ρ) determinations shown in Fig. 1 is the evident leveling off of the observed effective (mass) density to a value at the lowest temperatures considerably greater than the mass density of the helium-3 present. In fact, the mass density observed using this method is apparently approaching the same value of *twice* the mass density of contained helium-3 earlier indicated indirectly by King and Fairbank[3] from second sound velocity measurements and predicted independently by Feynman.[4]

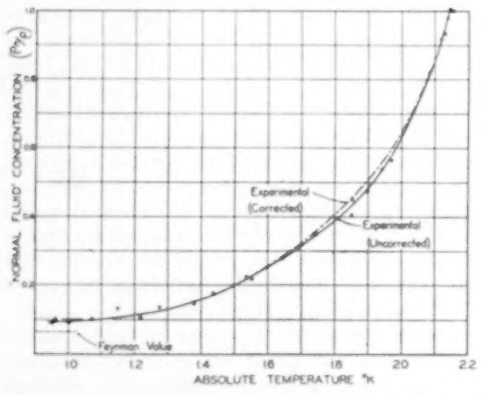


FIG. 1. "Normal fluid" concentration (ρ_n/ρ) for 3.3% (by mass) liquid helium-3–helium-4 mixture *versus* temperature (°K). Basic Andronikashvilli disk measurements, normalized for "ideal" (no correction) system to density of liquid mixture just above λ point. Circles (○) represent data taken with 5.220-sec period (in vacuum) system, triangles (Δ) with 3.589 sec system. Solid line ——— through these points represents uncorrected values; broken line — — — represents corrected dependence on temperature. Feynman limiting value indicated by dotted line - - - - -.

In Feynman's theory the motion of a helium-3 atom through a background of helium-4 atoms in the lowest energy state (superfluid) may be treated as a microscopic hydrodynamic problem. The helium-4 "backflow" about the helium-3 "obstacle" contributes the additional effective mass. At the lowest temperatures reached in the present direct investigation, the disks were thus essentially towing helium-3 atoms through a superfluid background, gaining the additional momentum of the helium-3 plus that of the participating "superfluid" (in the presence, of course, of some remaining rotons).

Although the coefficient of viscosity (η) for the mixture is not readily evaluated by this method, semiquantitative results show a distinct decrease with temperature. For example, the skin depth at 1°K is about that for the λ point, indicating a decrease in (η) roughly proportional to that in (ρ_n). Present measurements indicate a value of 4 or 5 micropoise at 1°K.

In a paper now in preparation the complete density data will be given in corrected form, plus the associated damping measurements and a discussion of their application to the corrections. Semiquantitative evaluation of the viscosity for this mixture as a function of temperature will be given.

* Supported in part by the National Science Foundation.
[1] E. L. Andronikashvilli, J. Phys. (U.S.S.R.) **10**, 201 (1946).
[2] Obtained from Isotopes Division, Oak Ridge National Laboratory, U. S. Atomic Energy Commission.
[3] J. C. King and H. A. Fairbank, Phys. Rev. **93**, 21 (1953).
[4] R. P. Feynman, Phys. Rev. **91**, 262 (1954).

Impedance Measurements on CdS Crystals*

HARTMUT KALLMANN, *Department of Physics, New York University, New York, New York,*
BERNARD KRAMER, *Department of Physics, New York University, New York, New York, and Physics Department, Hunter College, Bronx, New York*
AND
PETER MARK, *Department of Physics, New York University, New York, New York*
(Received June 28, 1955)

PREVIOUS measurements of changes in the impedance of ZnCdS luminescent powders due to excitation by light and high-energy radiation[1] has been extended to CdS crystals. These crystals were grown by the Standard Piezo Company of Carlisle, Penn-

sylvania using the Frerichs-Warminsky method.[2,3] The crystals are about 2 cm long, 0.5-1 cm wide, and several tenths of a millimeter thick; the electrodes are transparent gold layers evaporated on the flat sides of the crystals.

The values of the measured capacitance (C) and dissipation factor (D) were obtained using a General Radio 650-A impedance bridge (1000 cps). The results of measurements on 8 crystals are given in Table I. The first column gives the dark value; the second the value under visible illumination (tungsten source, Corning 4–96 filter, $\simeq$100 microwatts/cm^2); the third under visible illumination plus infrared irradiation (tungsten source, Corning 7–56 filter, 10 milliwatts/cm^2); and the fourth with the visible illumination only but using a bridge output of 12 volts rather than the usual 0.05 volt.

The tremendous increase in capacitance from the dark value is seen by comparing columns 1 and 2; if the system is looked upon as a parallel capacitance-resistance circuit consisting of the constant capacity C_0 (dark capacity of the sample) and a resistance changing with excitation, the measured capacitance would increase as $C_m = C_0(1+D^2)$. This is not the case, the increase is much larger than according to this formula. The very high capacitance values obtained under visible excitation indicate that the bulk of the crystal becomes so highly conducting that it acts like a very low resistor in series with less conductive thin layers adjacent to the electrodes, which alone contribute effectively to the parallel capacitance-resistance circuit. Even these layers, however, are quite conducting, as can be calculated from the C and D values. For the most sensitive crystals these calculated resistances are of the order of 10^5 ohms at high light intensities (large D), and agree with the measured dc photoconductivity. Using the measured C and D values, the thicknesses of these layers are found to be about 10^{-6} cm for the most sensitive crystal. Such layers have also been investigated in semiconductor work.

When infrared irradiation is added to the visible excitation, a quenching effect (decrease in capacitance change) is noted for all crystals except X-4. With this crystal, quenching is observed only at lower levels of light excitation. In column 4 it is seen that the increased ac field applied to the crystal also acts as a quenching agent except for crystal X-16 which shows a larger capacitance increase with the higher potential; perhaps in this case some barrier layers have been broken down by the higher field.

The capacitance change dependence on wavelength of one of these crystals (X-1) at various intensities is shown in Fig. 1. The capacitance change is a maximum around 5000 A, which is close to the absorption edge and corresponds to the peak observed in dc photoconductivity.[4] The large capacitance change found even at 4000 A, where the incident energy is completely absorbed in a small fraction of the total thickness, indicates that this layer reradiates some of the absorbed energy, thus exciting the remainder of the crystal by its own fluorescent light. The capacitance change increases roughly proportionately to the intensity; this is in marked contrast to the behavior with powders where the capacitance change increases roughly as log (intensity) tending to a saturation value[1] which is of the same order of magnitude as the dark capacity.

These results support the idea that the much smaller capacitance changes observed in highly photoconductive powders are due to their grain structure (high resistance between grains), and further they indicate clearly that the increased capacitance of these crystals is due to the large increase in the conductivity of the sample, since it is difficult to imagine any polarization due to trapped electrons or activators increasing the dielectric constant by a factor of up to 10^6. They also show that the simple model containing a fixed capacitor shunted by a light-sensitive resistance cannot explain the results; actually with increasing excitation the effective capacitive layers seem to shrink to small fractions of the initial thickness.

In addition the effects of fields on conductivity is appreciable. In almost all cases an increased field results in a marked decrease in conductivity; in one case an increase was noted.

Finally, the negligible effect of filled traps on the capacitance change can be pointed out. Upon cessation of excitation the capacitance drops quickly (within

TABLE I. Capacitance[a] and dissipation values under various conditions of illumination.

Crystal	Dark		Visible		Visible + infrared		Visible + 12 volt ac	
	C	D	C	D	C	D	C	D
X-1	12	0.02	2.0×10^4	5.0	1.7×10^4	4.9	1.9×10^3	2.1
X-3	17	0.02	405	0.3	79	1.0	57	0.2
X-4	34	0.03	8×10^3	1.1	8×10^3	1.5	3.3×10^3	1.4
X-8	19	0.02	146	0.36	45	0.43	34	0.16
X-13	25	0.03	7.1×10^3	2.3	6.6×10^3	1.9	...	...
X-16	7	0.06	105	0.78	90	0.72	2×10^3	3.9
X-17	13	0.06	1.3×10^4	9	1.0×10^4	9	...	...
X-18	34	0.35	1.7×10^7	11	1.4×10^7	11	...	...

[a] Capacitance in micromicrofarads.

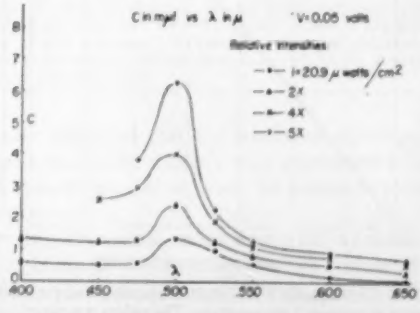


FIG. 1. Capacitance change as a function of wavelength at various intensities.

minutes) to the dark value although there is an appreciable number of traps which remain filled for many hours (as measured by succeeding rise curves).

* This investigation was supported by the Standard Piezo Company, Carlisle, Pennsylvania.

[1] Kallmann, Kramer, and Perlmutter, Phys. Rev. **89**, 700 (1953).
[2] R. Frerichs, Naturwiss. **33**, 281 (1946).
[3] R. Warminsky, PhD thesis, Technical University of Berlin, 1948 (unpublished).
[4] R. Frerichs, Phys. Rev. **72**, 594 (1947).

Magnetic Indications of Electronic Structure of the Conduction Band in Ge

J. H. Crawford, Jr., H. C. Schweinler, and D. K. Stevens
Oak Ridge National Laboratory, Oak Ridge, Tennessee
(Received June 27, 1955

CYCLOTRON resonance experiments[1,2] have firmly established that in the conduction band of Ge the energy surfaces are best represented by ellipsoids in k-space whose axes lie along the [111] directions. Moreover, since the minima of E *vs* k surfaces do not lie at $k=(0,0,0)$, this configuration leads to spatial degeneracy of states: eightfold if the minima lie within the Brillouin zone or fourfold if they lie at the zone boundaries. The magnetic susceptibility of charge carriers in semiconductors[3,4] yields information complementary to the results of cyclotron resonance. Moreover, such data provide a means of deciding whether four or eight ellipsoids best describe the states in the conduction band of Ge. In this letter, besides comparing the electron masses yielded by the two methods we wish to present tentative evidence for fourfold spatial degeneracy.

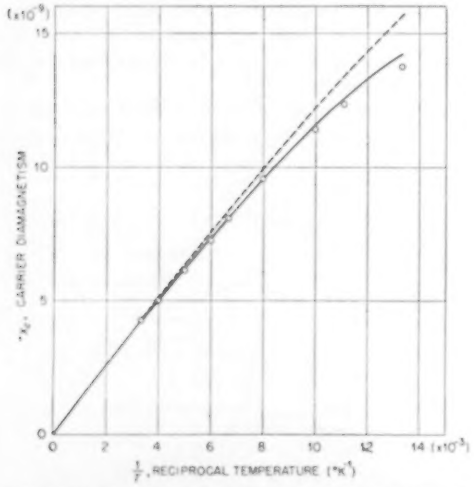


FIG. 1. Diamagnetic susceptibility resulting from conduction electrons *vs* reciprocal temperature. The points are experimental, the dashed curve calculated for $\omega_e=8$, the solid curve for $\omega_e=4$. In both cases $n=6.4\times10^{17}$ cm^{-3} (see text).

TABLE I. Electron concentrations and values of $\langle f_e^2\rangle$ obtained from the high temperature slopes of the χ_e *vs* $1/T$ curves for 5 n-type Ge specimens.

Sample	n(cm^{-3})	$\langle f_e^2\rangle$
1[a]	$\sim3\times10^{16}$	~50
2	7.3×10^{17}	48
3	9.0×10^{17}	44
4	3.6×10^{18}	36
5	9.0×10^{18}	19

[a] The electron concentration of this specimen was appreciably temperature dependent.

In the classical range of electron concentration, the susceptibility of extrinsic electrons in n-type Ge is given to first order by

$$\chi_e=n(\beta^2/3\rho kT)(3-\langle f^2\rangle), \tag{1}$$

where n is the electron concentration, β the Bohr magneton, ρ the density, and $\langle f^2\rangle$ is the square of the reciprocal mass-ratio tensor averaged appropriately for orbital motion in the magnetic field. For ellipsoidal surfaces it can be shown that $\langle f^2\rangle=(2M_t+M_l)/(3M_t^2M_l)$, where M_t and M_l are the transverse and longitudinal mass ratios $(M_i=m_i/m_0)$ whose values are 0.083 and 1.4 respectively.[1] These values yield $\langle f^2\rangle=54$. In Table I we list the values of $\langle f^2\rangle$ determined on several specimens from the temperature dependence of χ_e in the classical range. Values of n were obtained from Hall coefficient measurements.[5] The reliability of these $\langle f^2\rangle$ values is $\sim10\%$. Although agreement with the expected value of $\langle f^2\rangle$ is quite good at the lowest electron concentration, a significant decrease occurs with increasing n. Although this behavior is not yet completely understood, Herring[6] has predicted that a change in mass ratio might result if an impurity band is present which overlaps the conduction band at high donor concentration.

At high electron concentrations or low temperature, transition from classical to Fermi-Dirac statistics occurs. Under these conditions Eq. (1) becomes

$$\chi_e=(\beta^2/3\rho k)CT^{\frac{1}{2}}\omega_e(M^N)^{\frac{1}{2}}(3-\langle f_e^2\rangle)F_{\frac{1}{2}}'(\eta) \tag{2}$$

where $C=[4\pi(2m_0k)^{\frac{1}{2}}]/h^3$, ω_e is the number of ellipsoids, and M^N is the density of states mass given by $(M_t^2M_l)^{\frac{1}{3}}$. $F_{\frac{1}{2}}'(\eta)$ is the first derivative of the Fermi integral which can be obtained from the McDougall-Stoner tabulation[7] provided η or $F_{\frac{1}{2}}(\eta)$ is known. The value of $F_{\frac{1}{2}}(\eta)$ is obtained from n and the relation

$$n=C\omega_e(M^NT)^{\frac{1}{2}}F_{\frac{1}{2}}(\eta). \tag{3}$$

Hence by fitting χ_e *vs* $1/T$ curves in the transition range, it should be possible to determine whether ω_e is four or eight. There is a factor of two difference in the

density of states resulting from these possibilities, the larger value tending to shift the transition toward lower temperature for a given electron concentration.

The χ_a vs $1/T$ curve for sample 2 (Table I) is shown in Fig. 1. In this specimen n is essentially constant over the whole temperature range. The points are experimental and the curves are calculated from Eq. (3) for both values of ω_0. In order to make the comparison, the smaller value of $\langle f^2 \rangle$ indicated in Table I made it necessary to choose $n = 6.4 \times 10^{17}$ cm^{-3} as indicated by $\langle f^2 \rangle = 54$ rather than the value obtained from the Hall coefficient (7.3×10^{17} cm^{-3}). This modification should have little effect on the analysis. It is clearly evident from Fig. 1 that fourfold degeneracy gives the better fit to the data. There is reason to believe that the bending down of the experimental curve toward the highest $1/T$ values arises from higher order corrections in the magnetic susceptibility equation.

We are deeply indebted to Miss Louise Roth of Purdue University and C. S. Fuller of Bell Telephone Laboratories for providing the specimens used in these studies.

[1] Lax, Zeiger, and Dexter, Physica **20**, 818 (1954).
[2] Dresselhaus, Kip, and Kittel, Phys. Rev. **98**, 368 (1955).
[3] G. Busch and E. Mooser, Helv. Phys. Acta **26**, 611 (1953).
[4] D. K. Stevens and J. H. Crawford, Jr., Phys. Rev. **92**, 1065 (1953).
[5] The usual relation, $R = 3\pi/8ne$, was used in obtaining electron concentrations. We are indebted to J. W. Cleland of this Laboratory for the Hall coefficient measurements.
[6] Conyers Herring (private communication).
[7] J. McDougall and E. C. Stoner, Trans. Roy. Soc. (London) **A237**, 67 (1936).

Streamer Theory and the Value of α/p

K. Hasebe and K. Yamamoto
Nagoya University, Faculty of Engineering, Nagoya, Japan
(Received May 23, 1955)

THE streamer theory of the breakdown in gases was proposed by Meek[1] and, independently, by Raether[2] and is known to be applicable to relatively large values of $p\delta$ (p: gas pressure, δ: gap length). In the streamer theory the values of the breakdown fields in air are usually calculated using the values of α/p obtained by Sanders[3] (α: the number of electrons produced in the path of a single electron traveling a distance of 1 cm in the direction of the field). For a uniform field gap in air at atmospheric pressure, the difference between the values calculated by this theory and the observed values becomes greater as the gap length deviates from the value of 1 cm. For gaps shorter than 1 cm the calculated fields are higher than the observed values, one reason for which is that the breakdown at the shorter gaps is more likely to occur by Townsend mechanism rather than by streamer mechanism. For longer gaps the calculated fields fall below the measured values, for which effect there has as yet been no satisfactory explanation, although it has been discussed at length by Loeb and his associates, as well as by Meek, Raether, and Hopwood. Loeb and Meek pointed out that for longer gaps, the calculated ion density in the avalanche head is less than the critical value which is necessary to form the streamer. The field must therefore be increased to produce a higher ion density if a streamer is to be formed to cause the breakdown, in which case the breakdown develops through the formation of the mid-gap streamer.

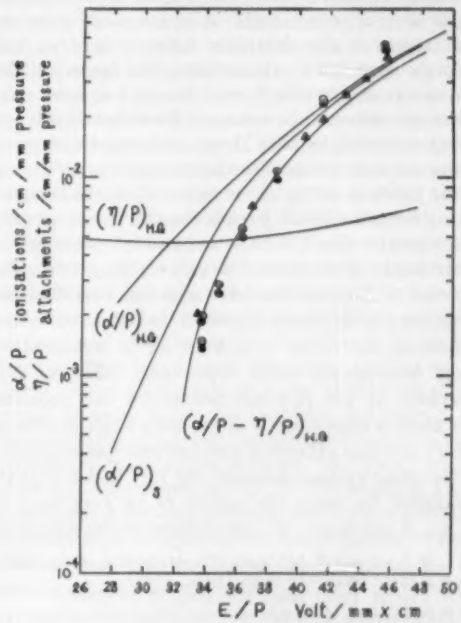


FIG. 1. Relations between E/p and the values of $(\alpha/p$, η/p, $\alpha/p - \eta/p)$ in air. $(\alpha/p)_{HG}$: the measured curve of Harrison and Geballe. $(\alpha/p)_S$: the measured curve of Sanders. △: the measured values of Paavola. $(\alpha/p)_P$. ⊙: the values of α/p obtained so as to give agreement between the measured fields and the calculated values of Meek's equation in air at atmospheric pressure $[E_r = E = 5.27 \times 10^{-7} \alpha\epsilon^{2.2}(x/p)^{-1}]$. ⊕: the values of α/p obtained so as to give agreement between the measured fields and the values calculated by Raether's equation in air at atmospheric pressure ($\alpha\delta = 17.7 + \ln\delta$).

Raether suggested that the conditions governing the propagation of the streamer may necessitate a higher field than is required to cause the avalanche-streamer transition.

In such discussions on the reason for the deviations, we entertain a doubt concerning the values of α/p used in the calculation. Sanders' data of α/p have been considered to be reliable, as they coincided with Masch's[4] data which were obtained by more accurate apparatus.

At that time, there were no adequate investigations on the effect of attachment upon the value of α/p.

Recently Harrison and Geballe[5] measured the values of α/p for air and other gases; they separated η/p from the real α/p, where η is the number of negative ions produced by the attachment to neutral molecules in the path of a single electron traveling a distance of 1 cm in the direction of the field. Former measured values of α/p for electronegative gases, such as air, may possibly have included the values of η/p.

Figure 1 shows $(\alpha/p)_{HG}$ and $(\eta/p)_{HG}$ obtained by Harrison and Geballe, compared with the formerly measured values of $(\alpha/p)_S$ by Sanders and $(\alpha/p)_P$ by Paavola.[6] Paavola's values are of great importance because they were measured at atmospheric pressure. Also shown is the difference between $(\alpha/p)_{HG}$ and $(\eta/p)_{HG}$ computed by the authors. The figure includes the values of α/p which must be used to give good agreement between the measured breakdown fields and the values calculated with Meek's or Raether's criterion. These values coincide with the curve of $(\alpha/p-\eta/p)_{HG}$ or the values of $(\alpha/p)_P$ in the region where the streamer theory is applicable (E/p is less than 38.5 v/cm-mm Hg or δ is greater than 2.5 cm at atmospheric pressure). In other words, if the values of $(\alpha/p-\eta/p)_{HG}$ or $(\alpha/p)_P$ are used in Meek's or Raether's equation, the calculated breakdown fields closely approach the measured values.

Because the values of α/p for air as measured by many investigators differ, more exact values of α/p, especially at low E/p are desired for this problem. The authors have recently developed a single avalanche theory as well as a theory of modified multiple avalanche. In the single avalanche theory, the very best results are obtainable by using the values of $(\alpha/p-\eta/p)_{HG}$ or $(\alpha/p)_P$. A description of these theories will be published.

[1] L. B. Loeb and J. M. Meek, *The Mechanism of the Electric Spark* (Oxford University Press, Oxford, 1941).
[2] H. Raether, Ergeb. exakt. Naturwiss. **22**, 73 (1949).
[3] F. H. Sanders, Phys. Rev. **44**, 1020 (1933).
[4] K. Masch, Arch. Elektrotech. **26**, 587 (1932).
[5] M. A. Harrison and R. Geballe, Phys. Rev. **91**, 1 (1953).
[6] M. Paavola, Arch. Elektrotech. **22**, 443 (1929).

Dependence of Light Amplification in Phosphors on Light Intensity

FRANK MATOSSI
United States Naval Ordnance Laboratory, White Oak, Maryland
(Received June 17, 1955)

CUSANO[1] has shown that ultraviolet excited luminescence in a ZnS:Mn, Cl phosphor film can be enhanced by the application of an electric field. This is interpreted[2,3] as a radiation-controlled electroluminescence ("photoelectroluminescence"[3]) inasmuch as the electrons excited by the radiation supply the electrons for impact excitation of the luminescence centers. Under the conditions of the experiment, the amplification depends on the radiation intensity such that the ratio $\rho=I/I_0$ (I=luminescence intensity with field, I_0=luminescence intensity without field) varies, for constant field, inversely as the square root of the excitation intensity. It will be shown that this square root law can be derived by an appropriate modification of a previous treatment[4] of the simultaneous effects of radiation and electric fields on phosphors.

In analogy to Eq. (2) of the paper mentioned,[4] we write

$$dN/dt=\eta-A_1Nm-A_2N(n-m+N)+\epsilon(m-N)+\gamma N, \quad (1)$$

$$dm/dt=\eta-A_1Nm+\gamma N. \quad (2)$$

Here N=number of conduction electrons; m=number of holes and empty centers; n=number of traps; η=number of electrons excited to the conduction band by incident ultraviolet radiation, used also as a measure of this radiation; A_1, A_2="transition probabilities" for recombinations and trappings respectively; ϵ and γ are constants for a given field strength, that disappear for zero field. The ϵ term describes emptying of traps by the field ($m-N$=number of filled traps). This is the essential process of "electrophotoluminescence" effects.[4] The γ term is added to take into account the excitation of electrons by impact with accelerated conduction electrons, which is the essential process of photoelectroluminescence. Quenching processes are neglected as immaterial for the present case, and the equations are written for dc fields.

The luminescence intensity is assumed to be $I=A_1Nm$. In equilibrium,

$$dN/dt=0, \quad dm/dt=0. \quad (3)$$

Then $I=\eta+\gamma N$, $I_0=\eta$ and

$$\rho=1+\gamma N/\eta. \quad (4)$$

N can be determined from Eq. (3) after elimination of m. This model is, of course, simplified (no separation of empty centers and holes; "bimolecular" recombination; thermal transitions left out). But the success of similar models in other respects justifies trying it for the present problem, although we cannot expect rigorous quantitative results.

For N a cubic equation is obtained, which reduces to a quadratic equation for certain approximative assumptions, for instance, if the ϵ effect can be neglected. In this case,

$$N=\tfrac{1}{2}(\gamma/A_1-n)+[\eta/A_1+\tfrac{1}{4}(\gamma/A_1-n)^2]^{\frac{1}{2}}. \quad (5)$$

This leads to the observed square root law if the conditions

$$(\gamma-A_1n)^2\ll 4\eta A_1\ll 4\gamma^2 \quad (6)$$

are satisfied. Then $N=(\eta/A_1)^{\frac{1}{2}}$ and $\rho\cong\gamma/(\eta A_1)^{\frac{1}{2}}\gg 1$. The conditions [Eq. (6)] state that the photoelectroluminescence process should prevail; that the radiation intensity η should not be too large; that a certain but not too large number of traps is necessary, depending on the efficiency γ of the photoelectroluminescence process. This is in agreement with the qualitative aspects of the observations and their interpretation.

Under other conditions, Eq. (5) leads to other functions $\rho=\rho(\eta)$, which can easily be derived. In many cases, ρ does not deviate much from unity. If the electrophotoluminescence effect is prevalent (γ small), it readily follows from Eq. (4) that $\rho\cong 1$.

[1] D. A. Cusano, Phys. Rev. **98**, 546 (1955).
[2] F. E. Williams, Phys. Rev. **98**, 547 (1955).
[3] F. E. Williams, Electrochemical Society Meeting in Cincinnati, May, 1955 (unpublished).
[4] F. Matossi, Phys. Rev. **94**, 1151 (1954).

Deformation-Induced Charge Flow in NaCl Crystals*

D. B. Fischbach and A. S. Nowick
Hammond Metallurgical Laboratory, Yale University, New Haven, Connecticut
(Received June 9, 1955)

When a single crystal of NaCl is plastically deformed in an inhomogeneous fashion, a transient electric current flows through the crystal, even when no external electric field is applied. This effect was observed when cleaved single crystal plates of NaCl approximately 1 mm×10 mm×10 mm, obtained from the Harshaw Chemical Company, were deformed in compression at room temperature. The deforming load was applied to a smaller area on one face of the specimen than on the other. A vacuum tube electrometer amplifier was used to measure the voltage developed across a 10^{11} ohm resistor connecting the two electrodes (which also served as press platens). The resistance of the crystal specimens at room temperature was of the order of 10^{14} ohms, as determined from the ionic conductivity.

The greatest charge flow occurs whenever a given stress is applied to the crystal for the first time; reapplication of the same stress gives much smaller effects, especially if flow had been allowed to go to completion on the first loading (as discussed below). The direction of flow is determined by the sense of the deforming stress *gradient;* negative charge flows in the external circuit away from the side of the crystal to which the higher stress is applied. This phenomenon was apparently first observed by Stepanow,[1] who did not, however, appreciate that the deformation must be inhomogeneous to cause a charge flow, and thus was not able to control the direction of the effect.

An example of the deformation-induced charge flow is shown in Fig. 1, where the total charge flow is about 3×10^{-11} coulomb. If the deforming load increment is removed before the current has decayed to zero, the residual current drops immediately to a value near zero, as shown in Fig. 2. It should be noted that only the increment is removed; the major part of the total load remains on the crystal. On reapplying the load increment, the charge flow is resumed. Successive cycles of removal and reapplication produce successively smaller effects, as shown in Fig. 2. Removal and reapplication of the load after the current has decayed to zero produces little effect.

This deformation-induced charge flow may be observed with or without colloidal graphite electrodes painted on the crystal, with the stress axis parallel as well as perpendicular to the plane of the electrodes, and with various geometries of inhomogeneous stress. Applied electric fields have little influence on the effect. Even under fields as large as 10^3 v/cm, the direction of the initial charge flow produced by deformation is determined by the sense of the stress gradient rather than by the direction of the applied electric field.

The observed current cannot be attributed to a polarization process, since all charge flow produced by a given deformation is in one direction, with no appreciable current in the reverse direction even on removing the total load from the crystal. The effect of the applied stress gradient must then be similar to that of an electric field, i.e., it produces actual electrolysis of the crystal. It seems reasonable to expect that charges are liberated in the initial deformation and are continuously separated by maintaining the load (to account for the decay with time shown in Fig. 1). The results of Fig. 2 show that it is the load *increment* and not the total load which operates to separate the charge carriers in time. This suggests that the charge carriers moved by the stress must be charged dislocations. Such a concept is not unreasonable. Seitz[2] has shown that a jog on an edge dislocation in NaCl possesses a charge of $\pm e/2$ depending on the sign of the last ion before the jog. Since a moving dislocation should generate vacancies

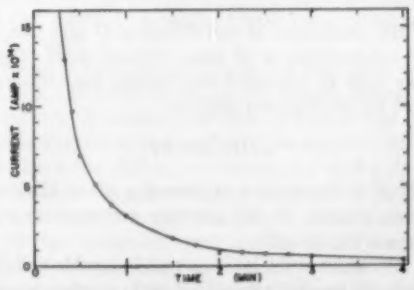


Fig. 1. Transient current produced by the inhomogeneous deformation of an NaCl crystal. At time equal zero, the load applied to the crystal was increased from 3 to 4 kg.

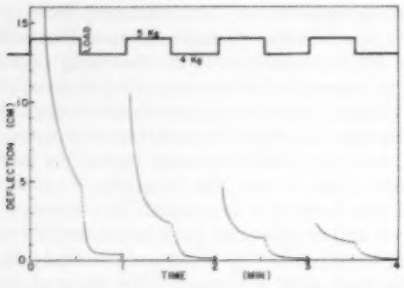


FIG. 2. The effect on the deformation-induced charge flow of removing and reapplying a load increment. The loading schedule is indicated by the heavy line at the top of the diagram. The current in amperes may be obtained by multiplying the deflection by 5×10^{-15}.

at jogs,[3] and since the energy of formation of a positive ion vacancy is probably less than that of a negative ion vacancy,[4] it follows that plastic flow should produce positively charged dislocations and leave behind in the lattice a net excess of (negatively charged) positive ion vacancies. When the deformation is inhomogeneous, it is then reasonable that the side on which the stress is concentrated develop a negative charge, as observed, since the positive dislocations move into the crystal, leaving behind negatively charged vacancies.

* This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

[1] A. W. Stepanow, Physik. Z. Sowjetunion **4**, 609 (1933).
[2] F. Seitz, Revs. Modern Phys. **23**, 328 (1951).
[3] F. Seitz, Phys. Rev. **79**, 1002 (1950); **80**, 239 (1950).
[4] N. F. Mott and M. J. Littleton, Trans. Faraday Soc. **34**, 485 (1938).

Self-Diffusion and Ionic Conduction in Crystals

ERNST KATZ

Department of Physics, University of Michigan, Ann Arbor, Michigan

(Received June 28, 1955)

THE coefficient of self-diffusion D and the ionic conductivity σ of ionic crystals, both referring to one type of transport mechanism, are customarily related by the Einstein relation

$$\sigma/D=Ne^2/kT, \tag{1}$$

where N is the number of ions of a given kind, e the electron charge, T the absolute temperature and k Boltzman's constant.

If the carrier of the ionic current is capable of forming electrically neutral aggregates with another imperfection (for example, with vacancies of opposite sign or divalent ions) then these neutral aggregates cannot contribute to the conductivity but they may contribute appreciably to the self-diffusion. In such cases deviations from (1) may be expected (and have been found) such that $\sigma/D<Ne^2/kT$.[1]

Recently a careful study on AgCl yielded an example where $\sigma/D>Ne^2/kT$.[2] The purpose of the present note is to point to a probable explanation of this type of deviation from (1). It is based on a correction which must be applied to the customary derivations of (1)[3] in the case of dielectrics.

The force which an external electric potential difference V, applied across a crystal of length l, exerts on a charge e inside a polarizable medium is not $eV/l=eE$, but greater. The effective field E^* and the field $E=V/l$ are related by

$$E^*=E(1\pm\Delta). \tag{2}$$

If the charge carrier is assumed to be a sphere with dielectric constant ϵ_1 in a homogeneous medium of dielectric constant ϵ_0, then

$$\Delta=\epsilon_1(\epsilon_0-1)/(2\epsilon_0+\epsilon_1). \tag{3}$$

Equation (2) leads straightforward to a generalized Einstein equation for dielectrics:

$$\sigma/D=Ne^2(1+\Delta)/kT>Ne^2/kT. \tag{4}$$

Equation (4) qualitatively explains the discrepancy observed by Compton and Maurer for AgCl. The conditions under which Eq. (3) is valid are not well satisfied for a conducting ionic vacancy or interstitial ion, so that (3) is only a rough approximation of the true value of Δ. Because the polarization on an atomic scale is not homogeneous and the path of the vacancy or interstitial ion avoids places where the polarization is high, it is reasonable to multiply the right side of (3) by a correction factor of the order $\frac{1}{2}$. We shall now discuss two typical cases: AgCl and NaBr. For interstitial ions the dielectric constant will be taken equal to the dielectric constant of the material since their contribution to the over-all dielectric constant is greatest by far. Thus one expects for interstitial ions for AgCl ($\epsilon=12$) $\Delta=1.85$ and for NaBr ($\epsilon=6$) $\Delta=0.84$.

For vacancies one may either take $\epsilon_1=1$, yielding for AgCl $\Delta=0.22$ and for NaBr $\Delta=0.19$, or one may make an estimate of Δ with $\epsilon_1=\epsilon_0$ by considering the motion of the ion jumping into the hole. This ion moves through the central part of a region with an extremely low polarization field exerted by other Ag^+ ions. Depending on which of several crude approximations one uses, he finds that this field is from 10 to 25 times as weak as that through which an interstitial ion moves, leading to $0.19>\Delta>0.08$ for AgCl and $0.1>\Delta>0.04$ for NaBr.

The essential conclusion of this crude analysis is that for vacancy transport Δ is so small that it may well have escaped detection but for interstitial transport, especially for ionic crystals with high dielectric

constant, Δ will be appreciable. This affords a new criterion for deciding whether electrolytic transport takes place interstitially or by vacancies. The fact that for NaBr Δ was found to vanish within the accuracy of the experiments[1] supports the vacancy transport assumption for this material. The fact that for AgCl Δ was found experimentally[2] to range between 1 and 2 is strong support for the interstitial transport assumption for the silver salt.

[1] See for example H. W. Schamp and E. Katz, Phys. Rev. **94**, 828 (1954).
[2] W. D. Compton, doctoral thesis under R. J. Maurer, University of Illinois, Urbana, Illinois, 1955 (unpublished).
[3] See for example N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, Oxford, 1940), p. 63.

Outgoing and Ingoing Waves in Final States and Bremsstrahlung*

HAAKON OLSEN†
Laboratory of Nuclear Studies, Cornell University, Ithaca, New York
(Received June 3, 1955)

IT is well known that, in calculating transition probabilities between stationary states, a scattered particle in the final state should be represented by a wave function whose asymptotic form is plane wave plus ingoing spherical wave, rather than by the usual scattering wave function. The importance of this point has been stressed repeatedly, and in particular in connection with the recent calculation of bremsstrahlung at high energies.[1]

On the other hand, it has been pointed out[2] that this particular choice of final state wave functions is only required if the differential cross section for a given direction of the scattered particle (final electron in the case of bremsstrahlung) is to be calculated. If it is asked for the cross section integrated over the direction of the scattered particle, then one can equally well use the outgoing wave type of wave function for the final state. This may be understood from the fact that in this case the direction of motion of the scattered particle merely represents a degeneracy of the final state. One may therefore use any complete set of states to represent the particle; in particular the outgoing wave modification may be chosen.

In the bremsstrahlung problem, these statements hold if the total cross section for emission of a quantum of a given energy and direction is desired. We need in this case a complete set of wave functions for the electron of a given energy. The statement that outgoing waves may be used for the electron will now be proved explicitly for a Dirac electron in a spherical symmetric potential.

Let $\psi_\pm{}^{(\mathbf{p},\mathbf{r})}$ denote the Dirac wave function with asymptotic form: plane wave with momentum $\mathbf{p}$ together with $\binom{\text{out}}{\text{in}}$-going spherical wave, μ being the spin label. In an arbitrary coordinate system this can clearly be written as

$$\psi_\pm{}^{(\mathbf{p},\mathbf{r})}=\sum_{jj_z lm} i^l\langle jj_z|lm\tfrac{1}{2}\mu\rangle \times Y_{l,m}{}^*(\mathbf{p}/p)e^{\pm i\delta_{l,j}}\Phi_{ljj_z}(\mathbf{r}), \quad (1)$$

where $Y_{l,m}$ are the spherical harmonics, $\langle jj_z|lm\frac{1}{2}\mu\rangle$ the Clebsch-Gordan coefficients, $\delta_{l,j}$ the phases and Φ_{ljj_z} the standing wave solutions

$$\Phi_{ljj_z}(\mathbf{r})=4\pi\left(\frac{E+m}{2E}\right)^{\frac{1}{2}}\begin{bmatrix} i\dfrac{p}{E+m}f_{l,j}(r)(\boldsymbol{\sigma}\mathbf{r}/r)\chi^{ljj_z} \\ g_{l,j}(r)\chi^{ljj_z}\end{bmatrix},$$

χ^{ljj_z} being the central field spinors,[3] and $f_{l,j}$ and $g_{l,j}$ the usual radial functions with asymptotic behavior:

$$f_{l,j}(r)\to(1/pr)\cos(pr-\tfrac{1}{2}l\pi+\delta_{l,j}),$$
$$g_{l,j}(r)\to(1/pr)\sin(pr-\tfrac{1}{2}l\pi+\delta_{l,j}).$$

The probability for a process described by a perturbation $\mathcal{O}$, integrated over angles and summed over spins in the final state, is proportional to

$$I_\pm=\sum_{\mu=\pm\frac{1}{2}}\int d\Omega_{\mathbf{p}_f}\left|\int d^3r(\psi_\pm{}^{(\mathbf{p}_f,\mathbf{r})})^\dagger\mathcal{O}\psi_i\right|^2. \quad (2)$$

On account of the orthogonality of the spherical harmonics and of the Clebsch-Gordan coefficients, the factor $\exp(\pm i\delta_{l,j})$ depending upon the choice of out- or ingoing wave modifications, disappears from the expression

$$I_\pm=\sum_{ljj_z}\left|\int d^3r\Phi_{ljj_z}^\dagger(\mathbf{r})\mathcal{O}\psi_i\right|^2, \quad (3)$$

giving the matrix element in terms of standing waves.

Turning now to the bremsstrahlung, it will be seen that the present remark introduces considerable simplifications:

Although it is necessary to use the final state wave function $\psi_-{}^f$ in order to obtain the correct differential cross section,[1] it is now clear that as soon as the cross section is integrated over angles of the electron, we can forget about the difference between $\psi_-{}^f$ and $\psi_+{}^f$. We may therefore take advantage of using $\psi_+{}^f$ in the final state, and hence we can deduce the bremsstrahlung cross section integrated over angles of the electron from the corresponding pair cross section by the familiar transformations $d\epsilon_1/k^2\to dk/kp_1$[2] and the change of sign of p_2 and ϵ_2. This conclusion holds whether the screening is absent, partial or complete. In particular

then, from the pair cross section formula, Eq. (43), of Davies, Bethe, and Maximon,[4] we thus find the bremsstrahlung spectrum at high energies:

$$(d\sigma/dk)=(d\sigma/dk)_{BH} -4a^2\frac{e^2}{\hbar c}\left(\frac{\hbar}{mc}\right)^2\frac{1}{kp_1^2}(\epsilon_1^2+\epsilon_2^2-\tfrac{2}{3}\epsilon_1\epsilon_2)f(Z). \quad (4)$$

Here $(d\sigma/dk)_{BH}$ is the Bethe-Heitler spectrum as given by Bethe,[5] including the effect of screening. Otherwise the notation is as in reference 4.

The Coulomb correction, which is given by the last term in (4), is therefore the same irrespective of the strength of the screening, and the relative correction is the same as for the pair cross section. We may furthermore use the same argument as was used by Davies, Bethe, and Maximon[4] in the case of pair production, to show that the form of the bremsstrahlung spectrum is essentially unchanged by the Coulomb correction.

In the case of no screening, the present result has in fact been obtained by Maximon[6] by integration of the differential bremsstrahlung cross section of Bethe and Maximon.[1] This, then, provides a good check on their calculation. On the other hand Bethe and Maximon[1] stated that there should be zero Coulomb correction to the Bethe-Heitler formula in the case of complete screening. This discrepancy between their result and ours arises from the fact that in bremsstrahlung the corrections due to screening and to the Coulomb effect cannot simply be superposed in the differential cross section. This is because the Coulomb correction occurs for small values of the momentum transfer q when the ingoing wave modification is used,[1] i.e., under the same conditions as the screening. But the customary expression for screening, *viz.*, the multiplication of the cross section by a form factor, is only valid in the Born approximation, i.e., only when the matrix element is *not* modified by the Coulomb correction. Therefore only the calculation of the cross section integrated over electron angles, with *outgoing* waves in the final state is reliable, and at present no result is available for the differential cross section of bremsstrahlung in the presence of both screening and Coulomb correction.

It is a pleasure to thank Professor H. A. Bethe and Dr. L. C. Maximon for interesting discussions.

* This work was supported by the Scientific Research Project OEEC 151.

† On leave of absence from Fysisk Institutt, Norges Tekniske Högskole, Trondheim, Norway.

[1] H. A. Bethe and L. C. Maximon, Phys. Rev. **93**, 768 (1954).

[2] Haakon Olsen, Kgl. Norske Videnskab. Selskabs, Forh. **28**, 10 (1955).

[3] Rose, Biedenharn, and Arfken, Phys. Rev. **85**, 5 (1955). Our expression (1) is in fact in the special case when $\mathbf{p}$ points in the direction of the z-axis, essentially equivalent to their expressions (33) and (34).

[4] Davies, Bethe, and Maximon, Phys. Rev. **93**, 788 (1954).

[5] H. A. Bethe, Cambridge Phil. Soc. **30**, 524 (1934).

[6] L. C. Maximon (private communication).

Preliminary Study of the Electrical Properties of a Semiconducting Diamond

JAMES J. BROPHY

Armour Research Foundation of Illinois Institute of Technology, Chicago, Illinois

(Received June 13, 1955)

THE Hall effect, electrical conductivity, and point contact rectification of a semiconducting diamond have been examined as the initial studies of electrical conduction in diamonds. The specimen investigated is a blue-white, rectangular parallelepiped 0.124×0.104 ×0.356 cm³, and is apparently a Type IIb diamond.[1,2]

The Hall effect was determined using a standard ac Hall apparatus operating at 27 cps and a dc magnetic field. Both indium solder and silver paint electrodes have been used, but neither type seems to be entirely ohmic at all temperatures. The specimen temperature was adjusted by immersion in a cooled or heated helium stream and the temperature determined by means of a thermocouple imbedded in the sample mounting cement. The temperature excursion in the present work was limited by contact instability at low temperatures and softening of the mounting cement at high temperatures.

The Hall constant is found to be independent of magnetic field up to the highest value employed, 10 000 oersteds, both at room temperature and at −40°C, the lowest temperature reached. The sample is p-type over the entire range. The logarithm of the Hall constant as a function of reciprocal temperature is found to be

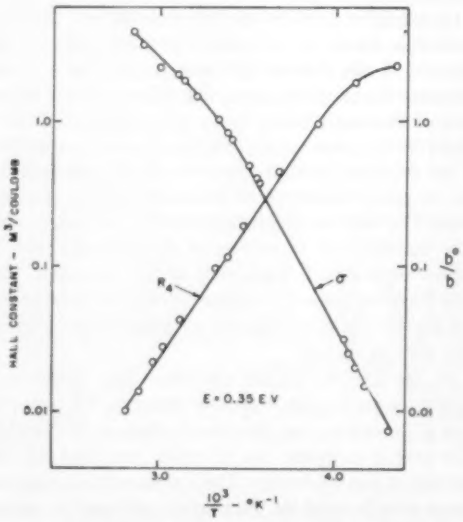


FIG. 1. Hall constant and conductivity as a function of reciprocal temperature for diamond sample.

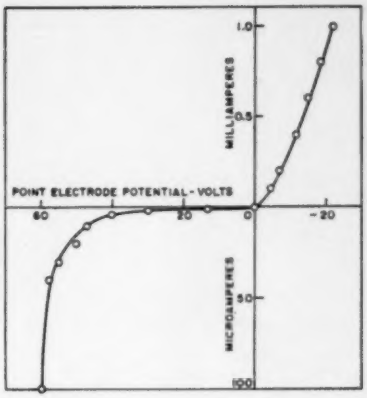


FIG. 2. Point contact rectification characteristic of diamond sample using tungsten probe.

linear except, perhaps, at the lowest temperatures, as shown in Fig. 1. The slope of the linear portion yields an activation energy of 0.35 electron volt. This agrees with the value reported by Leivo from resistance measurements.[3]

We attempted to investigate uniformity of the specimen by floating potential probe measurements using a tungsten point and a field of 3.0 volts/cm across the long dimension of the crystal. The results are inconclusive, showing variations in potential from point to point, which, however, are experimentally reproducible. It appears that there may be considerable conductivity inhomogeneity, on the surface at least. Point contact rectification examination showed that the entire surface was p-type.

Because of the suspected inhomogeneity, further measurements of the resistivity have been deferred. The conductivity of the specimen as a function of temperature was determined from resistance measurements between silver paint electrodes. These data are also shown in Fig. 1. We suspect that the curvature of this plot may be due to the electrodes, so that it seems clear that the major contribution to conductivity variation is thermal activation of carriers.

Although the nonuniformity of the specimen precludes determination of semiconductor parameters such as resistivity, carrier density, and carrier mobility, we have carried out calculations of these quantities based on the above measurements for general interest. We find the following results at room temperature:

$$p = 7 \times 10^{13} \text{ holes/cm}^3,$$

$$\rho = 760 \text{ ohm-cm},$$

$$\mu_p = 100 \text{ cm}^2/\text{volt-sec}.$$

These should be considered order-of-magnitude values only.

The point contact rectification characteristics have also been examined. A typical rectification curve for a tungsten point is shown in Fig. 2, although the exact shape depends somewhat upon the position of the probe on the surface. The sample will pass several tens of amperes in the forward direction with little difficulty, although the forward resistance is clearly appreciable. As mentioned above, p-type behavior is noted.

We hope to continue these investigations, particularly with regard to the inhomogeneities and the origin of donor sites.

[1] J. F. H. Custers, Physica **18**, 489 (1952).
[2] J. F. H. Custers, Physica **20**, 183 (1954). We are indebted to Dr. Custers of Diamond Research Laboratories for providing this specimen.
[3] W. J. Leivo and R. Smoluchowski, Phys. Rev. **98**, 1532 (A) (1955).

CONTENTS—continued

Letters to the Editor:

Structure of Transistor Current by [illegible] Emitter Radiation . . . . . J. Shields and W. B. Snyder 1325

Ripple Fields and the Rotation [illegible] . . . . . N. B. McCrea and J. C. Hancock 1326

Antiferroelectric Disk Resonators [illegible] . . . . . John R. Pellam 1327

Impedance Measurements on CdS [illegible] . . . . . Raymond Hofmann, Bernard Kramer, and Peter Mark 1328

Magnetic Ionization of Electrons [illegible] Conduction Band in Ge . . . . . J. R. Cuthbert, Jr., H. C. Schweinler, and D. K. Suescun 1330

Shammer Theory and the Value of α/ρ . . . . . K. Hanaki and K. Yamamoto 1331

Dependence of Light Amplification [illegible] Light Intensity . . . . . Frank Matossi 1332

Radiation-Induced Charge Flow [illegible] . . . . . D. B. Fischbach and A. S. Nowick 1333

Self-Diffusion and Ionic Conduction [illegible] . . . . . Ernst Katz 1334

Dumping and Ingoing Waves in Fluid [illegible] Stratification . . . . . Hashon Olsen 1335

Preliminary Study of the Electrical Properties of a Semiconducting Diamond . . . . . James J. Brophy 1335